Contributed Paper Sessions

CP 01: Statistical Modelling

Understanding the dynamic relationship between rainfall and temperature using multivariate time series models

Achal Lama and K.N. Singh
ICAR-Indian Agricultural Statistics Research Institute, New Delhi

Rainfall being a complex phenomenon governed by various meteorological parameters is difficult to model and forecast with high precision. Moreover, in recent times with changing climatic scenario the job has become more tough. For agrarian country like India which is largely monsoon dependent, rainfall modelling and forecasting takes centre stage. Hence, in this study we have made use of VAR model to understand the mean relationship between rainfall and minimum temperature. Further, to explore the dynamic relationship between the two series we have made use of dynamic conditional correlation a variant of MGARCH model. The study was carried out using monthly rainfall series of India and minimum temperature series from January, 1900 to December, 2006. This study revealed a positive and influencing relationship between minimum temperature and rainfall series. For comparative analysis, the performance of the multivariate model was compared with the univariate GARCH model in terms of different statistical measures such as RMSE, MAPE, etc. The modelling and forecasting efficiency of the multivariate model was recorded to be superior as compared to the univariate model to the tune of 5-10%. We believe our findings can help the policy makers and government for adequate planning for any adverse situation.

An Improved Beta Regression Model for Forewarning Helicoverpa armigera Infestation

Bishal Gurung[1], K.N. Singh[1] and Biwash Gurung[2]

[1]ICAR-Indian Agricultural Statistics Research Institute, New Delhi
[2]School of Agriculture, ITM University, Gwalior

Regression analysis is the commonly employed statistical tool for analyzing the “cause and effect” relationship. It has also been used to study how the weather variables affect the pest infestation data, which are sometimes measured on the proportional scale too. However, this technique may not be appropriate for statistical analysis of proportions as they are restricted to the interval (0, 1). As such, the assumptions of error terms being normal and

homoscedastic are violated. To this end, several researchers have identified different modeling strategies which could be employed to fit data that take values bounded on(0, 1).

The data collected under NICRA funded projects have data where % incidence is studied, taking weather variables as the independent variables. In such situations, the usual regression analysis is not appropriate. A reasonable alternative is to use transformations, but this again may lead to bias in estimation. Taking this into consideration, researchers have developed techniques to capture the changing proportional data such that interpretation is easier and also it becomes more flexible than transformation. In our research, we developed a statistical model for forewarning Helicoverpa armigera infestationby employing beta regression. By maximizing the likelihood function by employing the “optim” function through the analytical gradients available in R package the unknown coefficients are estimated.Moreover, Fisher scoring iteration through the use of expected information and analytical gradients were used to obtain better estimates through the use of “optim”. Residual diagnostics have also been carried out. The developed model could also be employed for forewarning pest infestation for other crops by using weather parameters as the independent variables.

Characterization and forecasting of Drought in Mahbubnagar, Telangana State

Amuktamalyada Gorlapalli[1], Supriya Kallakuri[2], Santosha Rathod[3], Meena Admala[2]

[1]College of Agriculture, Professor Jayashankar Telangana State Agricultural University, Aswaraopeta
[2]Professor Jayashankar Telangana State Agricultural University, Hyderabad
[3]ICAR-Indian Institute of Rice Research, Hyderabad

Drought is the period when there is no rainfall or low precipitation for prolonged time period. The whole world has witnessed drought. In Telangana, Mahbubnagar is one of the drought prone areas. So, it is important to know about the characterization of drought. This study is confined to development of drought indices using Standard Precipitation Index (SPI) for timescales 3 and 6 using popular time series models i.e., Auto regressive integrated moving averages (ARIMA), Artificial neural network (ANN) and Support vector regression (SVR) for forecasting of drought indices of SPI3 and SPI6 for Mahbubnagarof Telangana state. The drought indices classified based on the range of drought values, the results clearly showing that most of the months extremely and moderately drought. For this study monthly data of rainfall for the past 30 years i.e., January 1988 to May 2020 was collected from the Directorate of economics and statistics, Government of Telangana. The best forecasted model was selected by comparing the least values of MSE, RMSE and MAPE values of the

models. In this case SVR out performed ANN and ARIMA for both SPI3 and SPI6. The forecasting was done for 12 months i.e., June 2019 to May 2020.

Deep Learning Techniques for Forecasting Onion Market Prices

Kanchan Sinha and K.N. Singh
ICAR-Indian Agricultural Statistics Research Institute, New Delhi

Forecasting agricultural commodity prices is critical for farmers, governments, and agribusiness industries to make informed business decisions by managing risk and uncertainty. One of the most difficult aspects of onion price forecasting is its nature of high price volatility, production uncertainties, export policy, trader hoarding, poor market intelligence, inadequate storage and processing facilities, lack of robust demand and supply, and so on. By improving the accuracy of a suitable onion price forecasting model, policymakers and/or government bodies can make informed decisions based on market demand and supply. Although statistical time series models are highly interpretable, they struggle to achieve high precision in the complex agriculture scenario. In recent years, there has been a surge of interest in machine and deep learning techniques, which have gained prominence as a result of numerous applications in the fields of time series modelling and forecasting, as well as agriculture. There is a scarcity of research in the agricultural domain that is solely focused on onion price forecasting using a deep learning framework. This study contributes to addressing the problem of long-term historical reliance on onion market prices by analyzing and forecasting weekly prices of major consumer markets in India, namely, Delhi, Mumbai, and Bangalore, using recent advancements in deep learning models such as long short-term memory (LSTM) and its variants, attention mechanism based LSTM, convolutional neural network (CNN), hybrid CNN-LSTM, temporal convolution neural network, and temporal convolution neural network (TCN).Furthermore, they are also compared with one traditional statistical time series model viz., generalized autoregressive conditional heteroscedastic (GARCH) and two classic machine learning algorithms viz., time delay neural networks (TDNNs) and support vector regression (SVR). The findings show that deep learning models outperform statistical and machine learning models in terms of forecast accuracy measure criteria for all market, and they successfully address the onion price forecasting issue.

An improved zero-inflated count time series models for prediction of rice yellow stem borer population

Bojjareddy Nanda Kumar Reddy[1], Santosha Rathod[2], Supriya Kallakuri[2], Yerram Sridhar[2] and Pramit Pandit[3]

[1]Professor Jayashankar Telangana State Agricultural University, Hyderabad
[2]ICAR- Indian Institute of Rice Research (IIRR), Hyderabad
[3]Bidhan Chandra KrishiViswavidyalaya, Nadia

Count time series modeling is a popular statistical approach in which auto-correlated discrete count observations are considered as dependent variable, and the observations are assumed to be derived from Poisson or negative Binomial distributions. Crop pest modeling is one of the major areas of count time series modeling wherein daily or weekly counts of insects (pests) are considered as dependent variable of interest. As the pest population not occur regularly on daily basis there use to be many zeroes in weekly count data under such condition the classical count time series models may not yield better results alternatively zero excess count models can be used to model the data with excess zeroes.

In this study, the rice yellow stem borer (YSB) populations recorded using a light trap with an incandescent bulb along with weather parameters of major centres generated under All India Co-Ordinated Rice Improvement Project (AICRIP) from 2013-2021 were considered.

The residuals of the fitted models namely, Integer-valued Generalized Autoregressive Conditional Heteroscedastic (INGARCH), zero inflated Poisson autoregressive (ZIPAR), zero negative binomial autoregressive (ZINBAR) models were found significant for most of the centres. Therefore, to correct the classical model a two stage count time series methodology is proposed in this study. In first stage, the YSB populations are modelled using the count time series models and diagnostically tested using the multivariate Box-Pierce test. If this test is not significant, the procedure ends. In the second stage, when the residuals are significant, the BDS test is used to test for the nonlinearity of the residuals. If the residuals are found to be significant, the ANN model is fitted to the residuals to obtain the predicted values.

Finally, the ANN fitted residuals are combined with the count time series fitted values to obtain the final YSB population forecast. The proposed two stage methodology performed better than classical count time series models in both training and testing data sets, further, two stage zero inflated models outperformed all models which as it provides lowest root mean square error values. The proposed methodology for efficient early warning system to predict the YSB population could greatly contribute to the sustainable site-specific pest management strategies to avoid significant rice yield losses.

Spatio Temporal Neural Network with Application to Space-time Rainfall Forecasting

Mrinmoy Ray and K.N. Singh
ICAR-Indian Agricultural Statistics Research Institute, New Delhi

In an agriculturally reliant nation like India, accurate and effective forecasting methods for a variety of climate characteristics are vital. Rainfall is one of the most influential climatic elements on agricultural and allied sector production. In this context, rainfall forecasting is one of the most difficult tasks owing to the simultaneous existence of three patterns: temporal, spatial, and nonlinear. STARMA (space-time autoregressive moving average) models are frequently employed for modelling and predicting spatiotemporal time series data. However, the observed features of many space-time rainfall pattern include complex nonlinear dynamics, and modelling these patterns is usually beyond the scope of the conventional STARMA model. In addition, despite the widespread use of artificial neural networks (ANN) for modelling complicated nonlinear dynamics, they are incapable of handling spatial patterns. To overcome the aforementioned problem, in this article the Spatio-Temporal Neural Network (STNN) model has been proposed. The proposed approach has been empirically illustrated on annual precipitation data of six districts of northern part of West Bengal, India. The study reveals that proposed STNN has better modelling and forecasting precision over conventional STARMA.

Time Series Models Predicting Jassids and Thrips on Cotton at Akola of Maharashtra (India)

Satish Kumar Yadav[1], V.K. Rathaur[1], Ranjit Kumar Paul[2] and Prakash Kumar[2]

[1] Amity University, Noida

[2]ICAR-Indian Agricultural Statistics Research Institute, New Delhi

Autoregressive integrated moving average (ARIMA) and ARIMA with exogenous variable (ARIMAX) and integer-valued autoregressive (INAR) models were applied for prediction of infestation due to jassids (EmpoascakerriPruthi) and thrips (Scirtothripsdorsalis Hood & ThripspalmiKarny) occurring on cotton based on their population dynamics recorded overfive seasons (2005-09) of kharifat Akola district of Maharashtra State of India. While weekly per cent infestation due to jassids and thrips were used with ARIMA and INAR, the explanatory weather variables viz., maximum temperature (°C),minimum temperature (°C), relative humidity in morning and evening(%) and rainfall (mm)lagged by one week were considered for ARIMAX. Out of a total 92 observations, first 85 and the rest were used for estimation and forecasts of infestation levels of thrips and jassids. ARIMA and INAR models had a root mean square error (RMSE) of 2.03 and 1.60 for jassids and 0.91 and 0.30 for thrips, respectively for prediction of their mean infestations. RMSE values for forecasts of maximum infestations were marginally higher over that of mean infestations for both jassids and thrips. ARIMAX model indicated significance of morning relative humidity (MRH-1) and of both morning and evening RH (MRH-1&ERH-1) in determining mean and maximum jassid infestation,

respectively. While mean severity of thrips was influenced by minimum temperature (MINT-1) and MRH-1, maximum infestations were governed by MINT-1,MRH-1. Predictive performance of different models tested using mean square error (MSE) and root mean square error (RMSE) indicated that ARIMA and ARIMAX models did not differ significantly and the INAR model was to be better compared to ARIMA and ARIMAX models towards forecasts of both jassids and thrips on cotton at Akola (Maharashtra).

Performance evaluation of deep learning approaches for meteorological drought forecasting

Rajeev Ranjan Kumar and K.N. Singh

ICAR-Indian Agricultural Statistics Research Institute, New Delhi

Drought is an evolutionary process often triggered by precipitation deficiencies and may result in enormous economic losses. It is one of the many natural catastrophes that occur each year in various regions of the world. The prediction and early warning signs of drought are particularly important in the management and planning of agricultural resources before the onset of drought. Drought forecasting is one of the most arduous tasks owing to the simultaneous existence of complex nonlinear pattern. Data-driven artificial intelligence forecasting techniques are potential drought forecasting methods because they need less development time, fewer inputs, and are less complex than dynamic or physical models. This study explores the potential of the Deep Learning (DL) approaches for meteorological drought forecasting. The study reveals that, long short term memory (LSTM) outperformed than all the others competitive models in terms of forecasting accuracy.

Machine Learning based Approaches for Modeling Sub-divisional Rainfall in India

Himadri Shekhar Roy
ICAR-Indian Agricultural Statistics Research Institute, New Delhi

Agriculture in India remains vulnerable to weather fluctuations, sensitive to climatic changes and the looming threat of climate change may exacerbate this vulnerability. Mainly crop yields are sensitive to rainfall and temperature, any change in either will affect crop yields. Understanding the spatial and temporal distribution and changing patterns of climatic variables is critical for natural resource planning and management. In order to better understand the variability pattern in rainfall pattern and forecast short- and long-term changes in rainfall pattern, time series analysis can be a very useful technique. However, the traditional theory-driven numerical weather prediction (NWP) methods confront a number of issues, including the need for expensive computational resources, a lack of understanding of physical principles, and challenges in extracting knowledge from the vast amount of observation data. Recent development of machine learning and deep learning techniques

help to predict the rainfall pattern in different sub-divisions of India by using the sub-divisional rainfall statistics from the years 1871 to 2021. The prediction performance of the Recurrence Neural Network (RNN), Long Short Term Memory (LSTM), Convolutional Neural Network (CNN) models and machine learning based models are compared and examined in this study. The forecasting accuracy of each model is also evaluated. It is anticipated that the findings of this study would improve rainfall forecasting across different sub-division of India.

CP 02: Sample Surveys and Applications

Ameliorated Shewhart type Mean Control Chart Using Quartile Based Variant of the Ranked Set Sampling Scheme

Immad A Shah, SA Mir, MS Pukhta and S Maqbool

SK University of Agricultural Sciences and Technology of Kashmir, Srinagar, J&K, India

Sample techniques play an important role in determining the efficiency of control charts. The study was designed to develop a new Shewhart-type $\bar{x}\bar{x}$ control chart to monitor processes using a newly developed cost-effective method of ranked set sampling namely Stratified Balanced Quartile Ranked Set Sampling (SBGQRSS). SBGQRSS is a recent sampling design proposed based on traditional sampling of ranking set samples. ARL is utilized as a performance measure to evaluate the efficiency of SBGQRSS $\bar{x}\bar{x}$ control chart and other considered SRS, ranked set sampling (RSS) and extreme ranked set sampling (ERSS) charts by using Monte Carlo simulations. In most simulation scenarios, the SBGQRSS control chart is the best in comparison to RSS and its extensions. An application to real forest data illustrates the proposed method, with an evident increase in the sensitivity of the SBGQRSS based Shewhart-type$\bar{x}\bar{x}$ chart compared to other control charts.

Estimation of Finite Population Proportion from Geo-Referenced Survey Data

Ankur Biswas
ICAR-Indian Agricultural Statistics Research Institute, New Delhi

In recent times, due to extensive use of technologies in data collection, survey data often contain geographical information like geographical location. Such information is expected to be useful in survey data analysis. However, the commonly used survey estimation methodologies often ignore such information. Some researchers have explored the use of spatial information in estimation of finite population parameters but they are mostly limited to continuous survey variables. In many real-life applications, the variables collected from surveys are binary in nature. When the variable of interest is binary, it is appropriate to use a logistic model instead of the standard linear model. When sample is drawn using a complex

survey design, the simple logistic regression model may not be suitable. Under this study, a new estimator for population proportion has been proposed considering spatial logistic model for geo-referenced binary survey data. The spatial logistic model based on location specific Geographically Weighted Regression (GWR) model at the population level has been considered for the binary study variable. Expressions of variance and estimator of variance of the proposed estimator have been obtained using Taylor Linearization Technique. The performance of the proposed estimator was evaluated based on a spatial simulation study. Simulation results suggest that the proposed estimator outperforms usual Horvitz-Thompson and regression estimator of the population proportion.

Estimation of Energy Use for Onion Production in Karnataka State of India in the Purview of Sustainable Development Goal 7.0

Kaustav Aditya[1], Bharti[1], Maharani Din[2] and A. Carolin Rathinakumari[3]

[1]ICAR-Indian Agricultural Statistics Research Institute, New Delhi

[2]ICAR-Central Institute of Agricultural Engineering, Bhopal

[3]ICAR-Indian Institute of Horticultural Research, Bengaluru

Global attention has recently been focused on global warming and its negative impact on the livelihood of human population. One of the major contributors of global warming is excessive use of fossil fuel in agriculture, industrial sectors and transportation. As we know, India is an agriculture-based economy and after green revolution, the use of fossil fuel got aggravated in order to increase the food production to cope with the increasing population. Further, the drive for ever-increasing food production has resulted in extensive use of machinery, fertilizers, pesticides, and other natural resources and this resulted in excessive use of energy. Hence, to implement the Sustainable Development Goal (SDG) 7.0 of United Nations we need to estimate the amount of energy consumed for a crop throughout an agriculture year so that implementation of renewable energy sources as a substitute of the same can be planned. In addition, efficient energy use will reduce environmental issues and support sustainable agriculture. In the view of above, a survey on energy use in production of onion was carried out in Karnataka state of India. A stratified multistage random sampling design was used for the survey. Details of basic energy inputs and outputs, detailed energy consumption were collected and analyzed as per sampling methodology. It was found through the study, that there is a huge scope for energy conservation as well as implementation of renewable energy systems in the onion crop in the state of Karnataka.

Development of Survey Weighted Composite Indices under Complex Surveys

Deepak Singh[1], Pradip Basak[2], Raju Kumar[1] and Tauqueer Ahmad[1]

[1]ICAR-Indian Agricultural Statistics Research Institute, New Delhi
[2]Uttar Banga Krishi Viswavidyalaya, Cooch Behar

An index is constructed by a mathematical model representing multi-dimensional variables into a single value. Normally the indices are based on simple average/ weighted average method which does not consider the effect of multicollinearity among the indicator variables that are used for index construction except principal component analysis (PCA) based index method. However, the existing PCA-based method of index construction is incapable to incorporate the survey weights available in the complex survey which leads to an erroneous inference. The present study proposes some new methods of index construction for survey data that exclude the effect of multicollinearity as well as incorporate the survey weights and auxiliary information. In simulation study, It is found that the proposed methods of index construction are efficient in comparison to existing indices.

Multivariate Calibration Estimation using Nonlinear Constraints under Two Stage Sampling Design

Pathi Devendra Kumar, Kaustav Aditya, Tauqueer Ahmad and Ankur Biswas

[1]ICAR-Indian Agricultural Statistics Research Institute, New Delhi

The objective of sample surveys is to make inferences about a population from information contained in a sample selected from the same population. In survey sampling, the precision of the estimator is often increased by using the auxiliary variable. One such method of improving the precision of the estimates is calibration estimation. It was a well established fact that incorporating more than one auxiliary information in model assisted estimation improves the precision further than the case of single auxiliary information. The technique of incorporating multiple auxiliary variables in survey calibration is called multivariate calibration estimation. Now usually in calibration estimation the optimization is done using linear constraint function assuming independence of the auxiliary information while in real life conditions it is not always the case. Sometimes, the variables were found highly linearly related. To tackle such situations while using multivariate calibration a new strategy was proposed in this study. Nonlinear constraints based on two highly correlated auxiliary variables were used under two stage sampling design. The performance of the proposed estimators is evaluated through simulation studies. The empirical result shows that the developed estimators perform better than the existing estimators under two stage sampling design.

Survey Weighted Propensity Score Method for Impact Assessment

Raju Kumar,Deepak Singh and Tauqueer Ahmad

ICAR-Indian Agricultural Statistics Research Institute, New Delhi

Propensity score methods are commonly used to estimate the causal impact of a treatment or intervention when random assignment is not possible. The methods are particularly useful in situations in which it is unethical or impossible to randomly assign an intervention. Propensity score analysis (PSA) is now an exceedingly common statistical approach for estimating treatment effects from observational data. Propensity score matching (PSM) refers to the pairing of treatment and control units with similar values on the propensity score, and possibly other covariates, and the discarding of all unmatched units. Survey samples obtained using a complex survey design offer researchers the unique ability to estimate effects that are generalizable to the target population. Ignoring survey weights may lead to results that are not generalizable to the survey target population. Some researchers like Zanuto (2006), Dugoff (2014)etc. tried to incorporate survey weights from complex survey data in the propensity score analysis. In this paper, an attempt has been made to develop a new methodology of propensity score analysis to incorporate survey weights from complex survey data. The developed statistic has been illustrated with a real-life data set. It is observed through analysis that ignoring the survey weights affects the estimate of the population level.

Prediction of Urban Unemployment Rate in India using Grey Model

Pradip Basak, Kanchan Sinha and Mrinmoy Ray

[1]Uttar Banga Krishi Viswavidyalaya, Cooch Behar
[2]ICAR-Indian Agricultural Statistics Research Institute, New Delhi

Urban Unemployment Rate (UR) is a crucial indicator representing the livelihood of people in India. Accurate forecasting of the UR is essential for early identification of the socio-economic problems so that timely and targeted intervention, and proper policy planning can be done to reduce the same. Time series methodology utilized so far for the forecasting of Urban UR require monthly or quarterly data of sufficient length. In India, the quarterly estimates of urban UR in the Current Weekly Status (CWS) obtained through Periodic Labour Force Survey (PLFS) are available from the quarter April-June 2018 to July-September 2020 at the state as well as national level. Therefore, the usual methods of forecasting of UR may not yield reliable forecast in this type of small time series as the assumption on the data requirement will be violated. As a superiority to conventional statistical models, grey models require very limited data to build a forecast model (Deng, 1989). In this study, application of grey model has been considered on the quarterly estimates of urban UR for forecasting the unemployment in urban India. The Grey model shows excellent performance in forecasting the urban UR at the national level and at the state level, it shows good performances for most of the states.

CP 03: Inference and Multivariate/Methods

Characterization of Variability of Soil Zinc and Copper using Geostatistical Approach (Muzaffarpur, Bihar)

Abhineet Kumar, Nidhi and Mahesh Kumar
Dr. Rajendra Prasad Central Agricultural University, Pusa, Samastipur

Crop production is governed by three factors, viz., soil, inputs and climatic condition. Climatic conditions are beyond human control whereas soil and other inputs can be altered to enhance crop production. An understanding of soil properties is required for sustainable management of soil to reduce soil erosion and enhance soil health and fertility level. The variability of soil properties is needed for agricultural productivity, food safety and environmental modelling. The present study was conducted in soils of Muzaffarpur district in Bihar, India to understand the Variability of soil Properties Using Geostatistical Approach. zinc (Zn), copper (Cu)was measured. Spatial maps of soil properties were prepared using the semivariogram model through Kriging techniques. The value for nugget/sill of Cu was 0.25–0.75 indicating moderate spatial autocorrelation among the variables. Zinc (Zn) was highly concentrated in the study area. The present study suggests that the geostatistical model can directly reveal the spatial variability of soils.

Neural Network Modeling of Height Diameter Relationships for Himalayan Pine Through Back Propagation Approach

M. Iqbal Jeelani[1], Manish Kr Sharma[1], S.E.H. Rizvi[1], Afshan Tabassum[1], Fehim Jeelani[2] and Mansha Gul[1]

[1]Sher-e-Kashmir University of Agricultural Sciences and Technology, Jammu
[2]Sher-e-Kashmir University of Agricultural Sciences and Technology, Wadura
[1]Sher-e-Kashmir University of Agricultural Sciences and Technology, Jammu

Abstract: In this study, artificial neural network (ANN) models and nine conventional height diameter equations were employed to validate the height diameter relationship in Chir Pine plantations. 150 sample plots in the region of Jammu were used to gather the data. For the purpose of developing and validating models, the data was randomly partitioned into training and testing sets. Almost all the fitted height diameter models resulted in significant coefficients, which indicated that these models were able to capture the height and diameter

relationships. Out of the nine traditional height diameter (HD) models, Manfred HD model had the highest fitting precision, with lower values for Akaike Information criteria (AIC), Bayesian information criteria (BIC), Root mean square error (RMSE), Mean absolute error (MAE) and higher values for Coefficient of determination (R2) and Adjusted coefficient of determination (Adj R2). However, under cross validation, artificial neural networks (ANN) utilizing back propagation approach outperformed conventional models in every aspect as they resulted in lower values of Prediction error rates (PER) and other selection criteria. The flexibility and ability of ANN models to automatically find the hidden data patterns in the underlying data allowed them to surpass the other nine conventional height-diameter equations, according to our results, ANN models came out be superior in comparison to traditional modelling approaches.

Identification of Probable Potential Fishing Zones: A Case Study on Commercially Important Marine Resources in the Southern Coast of India

Eldho Varghese, J. Jayasankar, Reshma Gills and Shelton Padua
ICAR-Central Marine Fisheries Research Institute, Kochi

As exploratory fishery resource surveys are time consuming and expensive, landing centre-based surveys are frequently used in India to estimate the marine fishery resources. It is challenging to map the habitat of the resources in such surveys because the resources seen at the landing centre are not geo-tagged. Fishery management can benefit greatly from the use of geographic information system tools for habitat mapping, geo-referencing fish catch and fishing effort data, and linking catch to oceanographic and biochemical parameters. This is attainable if spatially scaled habitat information is provided for the resource. In this paper, an attempt has been made to identify the probable potential fishing zones of commercially important marine resources through a coordinated participative approach along with passive geo-referencing coupled with predictive modelling.

Growth Performance and Decomposition Analysis of Mungbean in Rajasthan

Rekha Rani, Uma Sah, Hemant Kumar, Jitendra Ojha and Mohit Katiyar
ICAR-Indian Institute of Pulses Research, Kanpur

Mungbean is third most important pulse crop after chickpea and pigeonpea in India. With the development of short duration varieties suitable for different agro climatic zones of India the crop has gained strides and popularity among farmers as it can be grown in all the seasons viz; kharif, rabi and summer. Present study was conducted to examine the growth performance and relative contribution of area and yield in increase in mungbean production in Rajasthan and its three potential districts viz; Nagaur, Jodhpur and Pali which together

contribute more than 50 % area in total mungbean cultivation in Rajasthan. The time series data were collected from 2000 to 2020 and bifurcated into two sub periods i.e. 2000-01 to 2009-10 (sub period I) and 2010-11 to 2019-20 (sub period II). The Compound Annual Growth Rate (CAGR), Cuddy Della Valle Index (CDVI) and Decomposition Analysis Model were used for analysis. The results revealed that during the entire period of twenty years the growth rate was positive in all area, production and productivity of mungbean at both state and district level. But the instability index was more than 30 in most of the cases which indicates the high variations in area, production and productivity of mungbean during the reported period of time. The sub period II can be declared as the best performing period as the period witnessed comparatively higher positive growth rate accompanied with lower instability index than the sub period I. Study further unveiled that the area effect was greater than the yield effect in most of the cases which indicates its greater share in increase in mungbean production at state and selected district levels. The study concluded that with the observed positive growth rate in area and its contribution in change in production, the state has scope to increase the area coverage under mungbean as 55 % districts of the Rajasthan are having less than 1 % area of total area under cultivation.

Hectareage Prediction Models for Paddy Crop of Middle Gujara

A.D. Kalola and R.R. Bhuva
B.A. College of Agriculture, Anand Agricultural University, Anand

The present investigation was undertaken with a view to identify the models for predicting the hectareage of paddy crop of the middle Gujarat region. The investigation was carried out on the basis of secondary data covering the period of nineteen years, (1998-99 to 2016-17). The District level data relating to hectareage, production, productivity and farm harvest prices of paddy were obtained from the published and compiled information by Directorate of Agriculture, Gujarat State, Gandhinagar. The linear multiple regression technique (basically Nerlovian type) was employed. The eight single equation and four simultaneous equation models were tried for paddy crop, the following models were selected on the basis of the values of adjusted coefficient of multiple determination. SE model-III for paddy is given below.

HEPD = 40960.532**** - 10.414*** HEBJ + 0.784 HEMZ - 1.187**** HEPDL + 3.720*** HEBJL + 5.588**** EYPD + 0.866 EYBJ - 6.205*** EYMZ - 6.833**** EPPD + 1.502 EPBJ (R2= 0.946)
HEBJ = 3261.298 - 0.061 HEPD + 0.108 HEMZ - 0.093 HEPDL + 0.337 HEBJL + 0.441 EYPD + 0.220 EYBJ - 0.619 EYMZ - 0.594 EPPD + 0.227 EPBJ (R2= 0.960)
HEMZ = 1816.343 + 0.028 HEPD + 0.147 HEBJ + 0.220 HEBJL + 0.649 HEMZL - 0.120 EYPD - 0.176 EYBJ - 0.092 EYMZ - 0.226 EPMZ - 0.106 EPBJ (R2= 0.850)

*, **, ***, **** Significant at the 20, 10, 5, 1 percent level of significance, respectively.

For the selected crops, SE model was recommended for prediction of the current hectareage on the basis of the adjusted coefficient of multiple determination ($\bar{R}\bar{R}$ 2). For Paddy

hectareage the main affecting factors viz., bajra hectareage, lagged hectareage of paddy, expected yield of maize and expected price of paddy. Expected yield and expected price of paddy were determining factors of bajra hectareage.

Use of weather variables for predicting whitefly count of cotton Crops through Generalized linear model approach

Manoj Kumar, Anil Jakhar, Anurag and Hemant Poonia

CCS Haryana Agricultural University, Hisar

In this study, an attempt was made to compare the Generalized linear model for the whitefly count data of cotton crop using weather variables. For the purpose weekly data of whitefly count (No. of adults/3 leaves) on three cotton varieties (Ganganagar Ageti, HS 6 and RS 2013) and one hybrid (RCH 650 BGII) grown in Cotton Section, CCS HAU, Hisar was recorded from 23 to 43 standard meteorological weeks during kharif season 2020 and 2021. The weekly meteorological data on rainfall, mean temperature, mean relative humidity, sunshine hours, wind speeds etc. for the same period were also considered for both the year. It was observed that variance is greater than mean for all four varieties so ordinary normal regression cannot be applied. The data was also tested for normality using Q-Q plot graph which shows that there is no normality in the data and data is over dispersed. It was observed that whitefly counts for all the varieties were positively skewed and shows that values of coefficient of skewness were 0.22, 0.014, 0.19, 0.30 for the year 2020 and for the year 2021 the value of coefficient of was found to be 0.71, 0.31, 0.45, 0.53. Histograms of whitefly counts over frequencies were also show no systematic pattern in occurrence of whitefly count. In most of the cases, occurrence of white fly scatters around its mean occurrence. It was found that occurrence of whitefly was very erratic. There is no linear relationship with time. It was observed that the weather variables i.e. minimum temperature, evaporation morning and evening, relative humidity morning and evening are significant with whitefly counts for all cotton varieties while wind speed variable is significant with only whitefly counts of HS 06 variety for the year 2020 whereas the sun shine, rainfall and rain days have no statistically significant correlation with whitefly counts data for all cotton varieties. The correlation coefficient for the whitefly count with weather variables for the year 2021 was observed that only relative humidity morning (Rhm) has positive and highly significant correlation with whitefly counts for all cotton varieties. After that whitefly counts were then

modelled as the function of significant variables using OLS, Poisson and negative binomial regression model for the both the year. It was observed that standard error of parameter estimates of negative binomial regression model showed low value as compared to the other studied model. When different models were compared on the basis of various goodness of fit criterial negative binomial model was found to be best with low AIC value and BIC value and high R2for all the studied cotton varieties.

Empirical Studies on Yield Distribution Modeling of Kinnow Crop

Geeta Verma, P.K. Mahajan, R.K. Gupta and Ashu Chandel
YSP UHF Nauni-Solan

In Agriculture, there is an extensive use of statistical probability distribution to model the crop yield distribution. In present paper, a field experiments were conducted at Indpur (location 1) and Indora blocks (location 2) of Kangra district of Himachal Pradesh and different probability distribution functions were fitted for prediction of number of trees in kinnow yield classes. The performance of distributions was evaluated using Kolmogorov-Smirinov test as the Goodness-of-Fit test. The result showed that Log-logistic (3P) and Beta (β) distribution were fitted best to predict number of trees in different kinnow yield classes for location -1 and location -2, respectively. These results can be useful for making better on- farm decisions that requires for prediction of crop yield based on theoretical information efficiently and precisely.

CP 04: Design of Experiments

Augmented Partial Four-way Crosses

Mohd Harun[1], Cini Varghese[1], Seema Jaggi[2], Eldho Varghese[3] and Anindita Datta[1]

[1]ICAR-Indian Agricultural Statistics Research Institute, New Delhi
[2]Indian Council of Agricultural Research, New Delhi
[3]ICAR-Central Marine Fisheries Research Institute, Kochi

The purpose of this investigatory study is to develop efficient mating plans suitable for crop and animal experiments. Augmented partial four-way crosses (APFC) have been introduced. Under a restricted fixed effects model for tetra-allele crosses incorporating general combining abilities effects, methodology of constructing any general class of APFC plan has been described and, the information matrices related to general combining abilities have been derived. The expressions for degree of fractionation, inverted information matrices, variance factors, amount of information factors and efficiency factors have been derived. A series of such APFC plans has been obtained along with various characterizing properties. These

plans can be of great importance while conducting a breeding experiment where some lines need to be highly-replicated for more precise interline comparisons compared to others.

Comparison of Uniformity Trial Data with Experimental Data for Plot Technique

Damor Bhavika, V.B. Darji and D.R. Darji
BA College of Agriculture, AAU, Anand

The problem in deciding optimum plot size frequently arises because variability of experimental area is unknown, which is main component of experimental error. Several workers have determined optimum plot size in many crops. But information regarding comparison for the plot size based on CV per cent of uniformity trial data and experimental data are not available in literature. Conducting the uniformity trial is costlier. Hence, there is a need to find out alternative way to replace uniformity trials utilizing the data of field experiments conducted on different crops at various Research Stations. The present study was under taken to compare plot size based on CV per cent of uniformity trials data with experimental data. Primary data were collected through uniformity trials conducted for Wheat crop and for forage crops Maize and Sorghum at Main Forage Research Station, Anand. Two methods were used to find optimum plot size for conducting trials i.e. Maximum curvature method and Mathematical approach.

The results showed decrease in CV values with the increase in plot size. The decrease in CV was proportional to the area added to form large sized plots upto 5 units size, thereafter the decrease in CV was not proportional to the area added to form large sized plot. Thus 10 units plot size adjacent to 5 units size appeared optimum plot size for field experiments on Wheat crop. In case of forage crops optimum plot size was 10 units size (for forage – Maize and Sorghum). This optimum plot size obtained through eye judgement.

The comparison between uniformity trial data and experimental data of Wheat, Maize (fodder), Sorghum (fodder) revealed that the uniformity trial conducted at Anand gave 10 units size plot (10 sq. m) as optimum plot size for experimentation are within the ranges of 9-12 sq. m. area of plot (experimental data). The results for maize and sorghum crops showed that 3-6 sq. m area most suitable plot size for forage crops. The comparison of other crops data (uniformity trial data & experimental data) showed that the optimum plot sizes estimated from uniformity trials are within the range of plot size estimated from the experimental data. It is advocated that for field experiments on these crops be conducted using plot size of respective crop and advised that there is no need of conducting uniformity trial on these crops.

Application of Repeated Measure Analysis in Impact Assessment

Manoj Kumar and KP Singh
ICAR-Central Institute of Agricultural Engineering, Bhopal

The objective of the paper was to assess the impact of rototill drill used for sowing of wheat without primary tillage operation using difference in difference method (DID). The DID have been widely used in econometrics and recently, it became popular in the fields of medicine. DID study designs are quasi-experimental design that aim to evaluate interventions in absence of randomization. It estimates the difference in pre-post changes comparing an exposed group to an unexposed group. In the present study, a group of farmers was exposed with demonstration of rototill drill and no exposure of roto till drill were given to second group. Pre and post measurements on energy use efficiency were taken from each farmers of every group. The experimental data was analyzed adopting MIXED procedure with REPEATED statement in SAS 9.3. All the effects (exposure, time and exposure-time interaction) were found to be significant at alpha= 0.05. In addition, wilks' lambda statistics also rejected the hypothesis of no time effect (p value 0.02). It was observed that rototill drill was able to significantly increase (p=0.002) energy use efficiency by 6.31(±1.93) in comparison to unexposed group. There was no significant change in energy use efficiency in unexposed group (p=0.47).

Effect of Missing Observation(s) on Third-order Response Surface Designs

Hemavathi M.[1], Eldho Varghese[2,] Shashi Shekhar[3] and Seema Jaggi[4]
[1]Pandit Jawaharlal Nehru College of Agriculture and Research Institute, Puducherry
[2]ICAR-Central Marine Fisheries Research Institute, Kochi
[3]Banaras Hindu University, Varanasi
[4]Indian Council of Agricultural Research, New Delhi

Response surface design aids in investigating the relationship between input and response variables by choosing an appropriate polynomial model. When the experimenter encounters the significance of lack of fit of a fitted second-order model while establishing the relationship between the input and response variables, sequential third-order response surface designs are advantageous for such situations. In some experiments, it is quite possible that some observations may be lost, destroyed, or impossible to access owing to a lack of resources, or that a particular factor level combination may be unable to be tested, or that some points will be avoided because of their high cost. This study examines the robustness of third-order response surface design to missing observation(s) and found that the maximum loss in information occurs when the observation at the design points having higher radii away from the design centre is getting lost and has minimum efficiency.

CP 05: Statistical Genetics/Bioinformatics

Next Generation Sequencing Data Analysis: Big Data Perspective

D.C. Mishra, Sudhir Srivastava, Neeraj Budhlakoti, K.K. Chatrurvedi,
S.B. Lal and Anil Rai
ICAR-Indian Agricultural Statistics Research Institute, New Delhi

Advancement in next generation sequencing technologies and low cost of sequencing has led to explosion of genomic sequence data. Consequently the analysis of such large and voluminous data has become a tedious task for the analyst. Among sequence data analysis, genome assembly is most difficult task especially when genome size is large. In genome assembly, millions of short reads are stitched together to form a single sequence of genome. Genome assembly is highly computational intensive and sometime converted into a NP hard problem. In this talk, the computational complexity along with other challenges of the genome assembly has been highlighted and the approaches of big data analysis to overcome these challenges has been discussed. In particular, algorithms based on ReduceMap and Hadoop technologies in the view of genome assembly/mapping which provides distributed and parallelized data processing and analysis has been discussed. Current approaches to assembly are outlined and the various software packages using cloud/parallel computing algorithms available online are introduced.

Integrated Approach for Genomic Prediction tohandle diverse Genetic Architecture

Neeraj Budhlakoti, Dwijesh Chandra Mishra, Sayanti Guha Majumda, Sudhir Srivastava,
K.K. Chaturvedi and Anil Rai

ICAR-Indian Agricultural Statistics Research Institute, New Delhi

Genomic selection procedures have proven to be useful in estimating breeding values and predicting phenotype using genome-wide molecular marker information. A number of statistical approaches i.e. parametric, nonparametric, semi parametric and other machine learning based methods have been proposed to predict individual breeding values by modeling the relationship between individual genotype and phenotype. It has been observed that parametric methods only perform satisfactorily when the system under study has additive genetic architecture. However, some nonparametric methods have also been developed to capture non-additive (dominance & epistasis) variances, but these generally fail to capture additive variances. The idea behind this study is to identify the best suitable model from each of the parametric and nonparametric categories and to develop an integrated model which can capture the merits of both models. It was observed from the results of comparative study between parametric and nonparametric methods that under additive architecture, GBLUP performed quite well while the performance of SVM was found to be encouraging in the case

of non-additive architecture. Keeping these results in the mind, a robust model has been developed for genomic prediction which can handle additive and epistatic effects simultaneously by minimizing their error variance. The developed integrated model has been evaluated using the measures of predictive ability and error variance. Moreover, it has been observed that our proposed model is robust to diverse genetic architectures.

Hybrid Approach for Imputing Missing Values in Proteomics Expression Data

Sudhir Srivastava, Dwijesh Chandra Mishra, U.B. Angadi, K.K. Chaturvedi and Anil Rai
ICAR-Indian Agricultural Statistics Research Institute, New Delhi

Missing values and heterogeneity are the most common problems in proteomics studies which can confound the important findings. Missing values in proteomics datasets may arise due to several reasons such as random errors, bad samples, or when features are below the detection limit in specific samples, etc. Further, expression data are always prone to heterogeneity due to technical and/ or biological reasons. Missing values can complicate the analysis and may lead to invalid conclusions. Therefore, appropriate data processing procedures for handling missing values are required for subsequent data analysis. We have proposed a hybrid imputation approach that can handle both types of missing values, i.e., missing at random (MAR) and missing not at random (MNAR).MAR specific methods such as KNN, SVD, BPCA and PPCA can be applied after identification of missing values due to MAR. For remaining missing values (due to MNAR),a novel imputation approach was developed using Gibbs sampling combined with advanced penalized regression techniques such as LASSO, elastic net, and ridge regression. The performance of proposed approach was compared with the existing approaches and found that performance of proposed method was at par or better than the existing approaches.

Estimation of DIVA Rates using NSP-ELISA Test for FMD Virus Sero-Surveillance in India during the Year 2008-21

Samarendra Das[1], Jajati K. Mohapatra[2] and R.P. Singh[2]

[1]ICAR-Directorate of Foot and Mouth Disease, Arugul, Bhubaneswar
[2]International Centre for Foot and Mouth Disease, Arugul, Bhubaneswar

Foot and Mouth Disease (FMD) is a severe and highly contagious viral disease of cloven-hoofed animals that has a significant economic impact. Usually, serological surveys are used to identify the infected animals among the vaccinated animals for proper implementation of the FMD control program. This has been achieved through Differentiating Infected from Vaccinated Animals (DIVA) principle by exploiting the differential antibody responses between vaccinated and infected animals. In other words, during natural FMD viral infection, arrays of non-structural proteins (NSP) are produced that elicit anti-NSP antibodies, which is not the case in animals vaccinated with an inactivated vaccine. Since 2008, random sampling design was adopted in India for the sero-surveillance activities. Thus, we present a statistical methodology for the estimation of the DIVA positive rates at the state and national level. For instance, for 2021, our methodology provides an estimate of 18.27% DIVA positive rate with 95% confidence interval of (14%, 22.5%) and 4.25% margin of error. Based on the developed methodology, an Rsoftware package has been developed for the user and available athttps://github.com/sam-dfmd/FMDSeroSurv.Furthermore, we tried to establish the relationship between the national level estimates of DIVA positive rate with field outbreaks of FMD through structural equation modelling approach. Our results indicate that the NSP-ELISA results are significantly associated with field outbreaks of one and two year lags. Besides, we observe a downward trend in DIVA positive rates as well as in FMD outbreaks over the years, which indicates the effectiveness of various measures implemented under FMD control programmes initiated by the Government of India. The developed methodology and obtained results can guide the FMD researchers and policy makers to identify potential disease free zones with vaccination in India. Moreover, the developed method and software can be extended to sero-surveillance of other viral diseases.

Study on Microbial Diversity under Varied agricultural Field Conditions of Wheat Crop

Sarika Jaiswal[1], Bharti Aneja[2,] Jaisri Jagannadham[1],[3], Bharati Pandey[1], Rajender Singh Chhokar[2,] Subhash Chander Gill[2,] Om Parkash Ahlawat[2], Anuj Kumar[2], U.B. Angadi[1,] Anil Rai[1,] Ratan Tiwari[2,] Dinesh Kumar[1,4]

[1I]CAR-Indian Agricultural Statistics Research Institute, New Delhi
[2]ICAR-Indian Institute of Wheat and Barley Research, Karnal
[1,3]Ashoka University, Rajiv Gandhi Education City, Sonepat
[1,4]Central University of Haryana, Mahendergarh

Wheat being the important staple food crop, plays a significant role in nutritional security. A wide variety of microbial communities beneficial to plants and contributing to plant health and production are found in the rhizosphere. The wheat microbiome encompasses extensive variety of microbial species playing a key role in sustaining the physiology of the crop, nutrient uptake and biotic/ abiotic stress resilience. This study presents the wheat microbiome analysis under six different farm practices, namely, Organic (Org), Timely Sown (TS), Wheat after Pulse Crop (WAPC), Temperature-Controlled Phenotyping Facility (TCPF), Maize-Wheat cropping system (MW) and Residue Burnt Field (Bur), respectively using 16s rRNA sequencing methodology. The soil samples collected from either side of the wheat row were mixed to get final sample set for DNA extraction under each condition. After the data pre-

processing, microbial community analysis was performed, followed by functional analysis and annotation. Abundance of phylum Proteobacteria was observed followed by Acidobacteria, Actinobacteria and Gemmatimonadetes in majority of the samples while relative abundance was found to vary at the genus level. The CAZy analysis showed high number of glycoside hydrolases genes in TS, TCPF and WAPC samples, while Org, MW and Bur samples predominantly had glycosyltransferase genes and the lowest carbohydrate esterase. Also, the Org and TCPF samples showed lower-diversity while rare and abundant species ranged from 12-25% and 20-32% of the total bacterial species in all the sets, respectively. These variations indicate that the different cropping sequence had a significant impact on soil microbial diversity and community composition, which characterizes its economic and environmental values as a sustainable agricultural approach to maintaining food security and ecosystem health.

Whole Genome-based identification of non-coding RNAs in Black Pepper (Piper nigrum L.)

Baibhav Kumar, Bibek Saha, UB Angadi, Anil Rai, Mir Asif Iquebal

ICAR-Indian Agricultural Statistics Research Institute, New Delhi

Long non-coding RNAs (lncRNAs) and circular RNAs (circRNAs) are the two classes of non-coding RNAs (ncRNAs) present predominantly in plant cells and have various gene regulatory functions at pre and post-transcriptional levels. Previously deemed as “junk”, these ncRNAs have now been reported to be an important player in gene expression regulation, especially in stress conditions in many plant species. Black pepper, scientifically known as Piper nigrum L., despite being one of the most economically important spice crops, lacks studies related to these ncRNAs. From a panel of 53 RNA-Seq datasets of black pepper from 6 tissues, namely, flower, fruit, leaf, panicle, root and stem of 6 black pepper cultivars, covering 8 bioprojects across 4 countries, we identified and characterized a total of 6406 lncRNAs. Further downstream analysis inferred that these lncRNAs regulated 781 black pepper genes/ gene products via miRNA-lncRNA-mRNA network interactions, thus working as ceRNAs. The interactions may be various mechanisms like miRNA-mediated gene silencing or lncRNAs acting as endogenous target mimics (eTMs) of the miRNAs. 35 lncRNAs were also identified to be potential precursors of 94 miRNAs after being acted upon by endonucleases like Drosha and Dicer. Tissue-wise transcriptome analysis revealed 4621 circRNAs. Further miRNA-circRNA-mRNA network analysis showed 432 circRNAs combining with 619 miRNAs and competing for the binding sites on 744 mRNAs in different black pepper tissues. These findings can help researchers to get a better insight to the yield regulation and responses to stress in black pepper in endeavor of higher production and improved breeding programs in black pepper varieties.

Simultaneous Selection for Yield and Stability in Pigeonpea of North East Plain Zone of India

Hemant Kumar
ICAR-India Institute of Pulses Research, Kanpur

Yield has been given most priority for varietal development program in almost all the crop and stability of the yield becomes an important criterion in pigeonpea, as most of the pigeonpea being cultivated in rainfed condition. For varietal development program multi-environment trails are conducted to advance the trial at different level. Multi-environmental trials have generally significant genotype main effects and genotype x environment interaction (GEI) effect and, therefore different univariate and multivariate stability methods have been used to study the GEI. Eberhart and Russell model is used for assessment of cultivar response to environmental changes using a linear regression coefficient (bi) and the variance of the regression deviations (sd2). Higher the value of sd2 for a genotype lower its stability. Based on the Eberhart-Russel model, an stability index is proposed to rank the genotypes. This index is the weightage of stability and yield component and higher the index value better is the genotypes. The index of 15 promising pigeonpea genotypes were computed with two different weight of yield (50% and &80%) and stability component (50% and 20%). These genotypes were evaluated at locations viz. Warangal, Gulbarga, Tandur, Banglore, Lam and Berhampore representing the North East Plain Zone of All India Coordinated Research Project on pigeonpea program during 2020-21. Ranking of genotypes are done based on two different weight of stability and yield component.

Application of PCA and Factor Analysis for Selection of Cotton Genotypes using Biometrical Characters

D.J. Parmar, J.G. Babariya, J.M. Dholariya and M.P. Patel
B.A. College of Agriculture, AAU, Anand

The data on eleven different biometrical characters of thirty genotypes of cotton were collected from RRS, AAU, Anand during kharif season of the year 2020 for the present study The experiment was conducted in randomized complete block design with three replications. The data on biometrical characters days to flowering, days to 50% boll bursting, plant height, number of monopodia per plant, number of sympodia per plant, number of bolls per plant, boll weight, ginning percentage, seed index, lint index and seed cotton yield per plantwere subjected principal component analysis and factor analysis. The present study was carried out to study the multivariate techniques for selection of cotton genotypes and to assess the biometrical characters for selection of genotypes of cotton.Based on the principal component

analysis, first three PCs were retained out of eleven PCs, these three PCs explained 73.46% of variability. Seed cotton yield per plant had maximum contribution followed by days to 50% boll bursting, lint index, ginning percentage number of bolls per plant and days to flowering in PC1 and PC2. The traits number of bolls per plant, number of sympodia per plant, boll weight and seed cotton yield per plant were categorized in one group days to flowering, days to 50% boll bursting and seed index in the second group; number of monopodia per plant and lint index were categorized as third group as these trait vectors had positive correlation with each other since the angle between them was less than 90o. The results of biplot showed that performance of genotypes G23 and G16 were found high in plant height and seed cotton yield, respectively. Genotype G30 gave highest performance in days to 50% boll bursting, days to flowering and seed index, whereas G26 had highest number of monopodia per plant and lint index. The results of factor analysis revealed that the eleven characters were categorized into three groups based on the high factor loading. The first group contains seed cotton yield per plant, boll weight, number of bolls per plant and number of sympodia per plant and they were named as “yield attributing” group as they are major yield contributing characters. Number of monopodia per plant, days to flowering, days to 50% boll bursting and seed index characters fall under vegetative growth attributing group. Ginning percentage and lint index were named as quality attributing group.

RbpRnaDB: A database for RNA-binding proteins and their combinatorial interactions with miRNA to explain the miRNA biogenesis model

Prakash Kumar1, Upendra Kumar Pradhan1, and Ravi Shankar 2

1ICAR-Indian Agricultural Statistics Research Institute (ICAR-IASRI), New Delhi

2CSIR-Institute of Himalayan Bioresource Technology (CSIR-IHBT), Palampur

A database of RNA-Binding Proteins (RBP)-miRNA interaction studies describes several RBPs and associated causal networks that are important for miRNA biogenesis. This database has 3 major modules. 1) RNA-centric and protein-centric RBP-miRNA interaction plots (binding site of various RBPs on different primary, precursor, and mature miRNAs regions) and their database search module; 2) Association of miRNA with pairwise different combinations of RBPs search module; 3) A comprehensive visualization platform of miRNA-RBP causal association networks module. The data set used here by extensive investigation of a large amount of high-throughput sequencing data taken from ENCODE, GEO, and miRbiom web servers. The correlation combination of miRNA with two different RBPs was also implemented utilizing the MongoDB database, and each combination resulted from the interactive table linked with a plot for further study with the expression value from a different sample size of given conditions. A dynamic search module for different conditions and different sample size-based correlation calculations with

correlogram plotting is executed. All the miRNA-RBP and their associated protein networks were visualized utilizing JavaScript, JSON, and D3j in the miRNA-RBP-proteins association networks module section. The weight of each link was derived from Bayesian Network Analysis (BNA).

Publically available miRNAs classification using machine learning models

Bulbul Ahmed
G D Goenka University, Sohna, Gurugram

MicroRNAs (miRNAs) are collections of short nucleotides with a 21–25 base pair range that aid in the regulation of gene expression for a variety of purposes. Due to their short read length, these miRNAs are intricate and difficult to identify. Therefore, using more sophisticated computational methods may help to more accurately detect these sequences. Various computational techniques have been utilised in recent years to classify a variety of biological data. Publicly accessible miRNA sequences were employed in this study for binary classification, and a dictionary of nucleotide sequences with a fixed length of 22 nucleotides was used for numeric representation. Different machine learning techniques viz. Support Vector Machine, Random Forest, and Random Tree classifiers have been used and it is found that random forest classifier outperform compare to other.

	Random Forest Classifier	Random Tree Classifier
True Negatives: 54 False Positives: 53 False Negatives: 36 True Positives: 55 Accuracy 55.05%: Precision 50.93% Recall 60.44% F1 Score 552.76% F2 582.63% Specificity 50.47% MCC 0.11	True Negatives: 50 False Positives: 57 False Negatives: 24 True Positives: 67 Accuracy 59.09%: Precision 54.03% Recall 73.63% F1 Score 62.33% F2 68.65% Specificity 46.73% MCC 0.21	True Negatives: 56 False Positives: 51 False Negatives: 34 True Positives: 57 Accuracy 57.07%: Precision 52.78% Recall 62.64% F1 Score 57.29% F2 60.38% Specificity 52.34% MCC 0.15

Identification of major biogeochemical cycle regulatinggenes in river Ganga and Yamuna through metagenomic approaches

Ritwika Das
ICAR-Indian Agricultural Statistics Research Institute, New Delhi

Biogeochemical element cycling processes have major impacts on agriculture, climate change as well as overall development of humankind. Microorganisms play important roles in molecular transformation of these elements between available and unavailable forms through enzymatically catalysed and thermodynamically constrained redox reactions. Microbes in natural ecosystems exist in the form of complex and highly connected networks that share and compete for metabolites. Due to rapid advancements in sequencing technologies, now-a-days metagenomic approaches facilitate the analysis of these microbial processes directly in its native environment. Microbial communities in aquatic ecosystems regulate the food web dynamics as well as biogeochemical processes but aquatic microbial diversity is immense and a general challenge is to understand how metabolism and interactions of single organisms shape microbial community dynamics and ecosystem-scale biogeochemical transformations. Ganga and Yamuna are two major rivers of India which greatly affects the civilization, agriculture and industries of North India. Various beneficial microorganisms have been found in these 2 rivers through metagenomic data analyses. In this study, microbial genes regulating 3 major element cycles, i.e., carbon (C), nitrogen (N) and sulfur (S) present in Ganga and Yamuna have been identified and annotated through metagenomic approaches using publicly available datasets (Behera et al., 2020). Comparative analysis of these river metagenomes has also been performed in light of their biogeochemical cycling potentials.

CP 06: Informatics

Prediction of Biosynthetic Gene Clusters and Drug Discovery using Metagenomics

Mohammad Samir Farooqi, K.K. Chaturvedi, D.C. Mishra, Sudhir Srivastava,
S.B. Lal, Anu Sharma and Neeraj Budhlakoti

ICAR-Indian Agricultural Statistics Research Institute, New Delhi

Specialized metabolites (also called as secondary metabolites) derived from bacteria, fungi, marine organisms and plants constitute an important source of antibiotics, anti-cancer agents, insecticides, immunosuppressants and herbicides. There is a growing demand for novel approaches for the mining of the uncultivated biodiversity existing in challenging environments and competitive communities, as a potential source of biologically active compounds. Metagenomic analysis has brought a revolution in various fields of biological research, notably drug discovery..Metagenomics permits both the discovery of genes as well as elucidates the biosynthetic pathways of both cultivable and uncultivable organisms. Biosynthetic gene clusters comprise a group of physically co-localized genes that together encode enzymes for the biosynthesis of a specific metabolite. Developments in structural and functional metagenomics have paved the way to discover novel genes and metabolic pathways for obtaining bioactive compounds. Functional metagenomics can be considered as a potential tool for discovery of useful gene products from the 99% of the uncultured microbial world since it does not require prior knowledge about genes or their hosts. It has made possible to identify novel antibiotics, as well as proteins involved in antibiotic resistance. Several bioactive metabolites have been obtained from metagenomic samples with immense therapeutic potential.

Extreme Learning Machine for Big Data Analysis

Sumanta Kumar Das
ISSA, DRDO, India

Until a few decades ago, we could normally analyze data of manageable sizes (from a few kilo-bites to mega-bites) using statistical techniques. Now we are more familiar with Big Data (BD)(from a few giga-bytes to tera-bytes) that is detailed over all spectrum of data ranging from continuous to discrete or ordinal to nominal categories. With the advancement of extreme learning machine (ELM) technology along with the increase in computing power of machines and the unimaginable improvements in memory power have revolutionized the statistical way of analyzing data. This would not have been possible without a fundamental foundation statistical tool. We will discuss in this paper how ELM techniques are evolving to advanced ML techniques and are often being applied for big data analysis.

Analytics Dashboard for Landscape Diagnostic Survey Data under Cereal Systems Initiative for South Asia

Soumen Pal, Alka Arora, Sudeep, Ajit, S.N. Islam and Ranjt Kumar Paul

ICAR-Indian Agricultural Statistics Research Institute, New Delhi

As a part of Cereal Systems Initiative for South Asia(CSISA), in India, Landscape Diagnostic Survey (LDS) was made by the Krishi Vigyan Kendras (KVKs) under the guidance of International Maize and Wheat Improvement Center (CIMMYT)to collect data on rice, wheat and/or maize cropping system in farmers' field using Open Data Kit (ODK) platform. LDS has so far been deployed in eight Indian states covering more than 100districts for detailed diagnosis of current production practices and yields. The data, after curation, has been transferred and then hosted as DBMS in ICAR Data Centre situated at ICAR-Indian Agricultural Statistics Research Institute, New Delhi. A dynamic data-driven dashboard has been designed and developed to draw meaningful inference out of this LDS data. This dashboard provides analytics on important parameters which can influence on crop yield. To study the interaction effect of different factors on grain yield of a particular crop, provision has been made in the dashboard to incorporate multiple parameters simultaneously. This kind of study can further be drilled down to state, district, up to block level using the dashboard. The dashboard can help in identifying reason(s)for the gap in varietal adoption by the farmers, and thus can assist in finding proper agricultural extension policy accordingly. Some GIS-based maps, based on certain parameters, have been incorporated as a part of this dashboard for visual comprehension of the data. This study can help the KVKs, Agricultural Technology Application Research Institutes (ATARIs) and higher extension authorities to frame region-specific agricultural strategies in the country.

Aavishkar-an AI based Mobile App for Rice Pest Detection

B. Sailaja[1], Ch. Padmavathi[1], D. Krishnaveni[1], B. Sreedevi[1], Santosha Rathod[1], M Srinivas Prasad[1], R. Mahender Kumar[1], D.Subrahmanyam[1], S.Gayatri[1], R.Nagarjuna Kumar[2] and R.M. Sundaram[1]

[1]ICAR-Indian Institute of Rice Research, Rajendranagar, Hyderabad
[2]ICAR-Central Research Institute for Dryland Agriculture, Hyderabad

Rice is the most important crop in India. It is widely cultivated in diverse ecologies and agro ecological zones. Biotic stresses play a crucial role in limiting the rice production. Farmers often face critical situations for diagnosis and management of pests, diseases and weeds. Integrated pest management practices need to be employed harmoniously to meet the above challenges. This demands complex decision making in the light of vast information knowledge base. Artificial Intelligence (AI) and mobile based applications have opened up new challenges as well as opportunities to fulfil the needs of farmers for up-to-date and precise information. Mobile phones have been spreading fast among farmers for exchanging the information among each other. Hence, an intelligent mobile based application is needed to reach farmers easily and generate quick decisions for real time problems facing at the field level. An Android app was developed for collecting the basic information on crop, diagnosing the pest and generating control measures. About thousand images of major insect pests, diseases and weeds of rice crop were trained in the database. Machine/deep algorithms were used for image processing. The alpha version of the app is available in IIRR website. Validation of the app in farmers’ fields is in progress and the beta version of app will be uploaded in the Google play store.

Machine Learning Based Mapping And its Acreage Calculation for Rice Crop using Sentinel 2 in Cooch Behar District Of West Bengal, India

Chandan Kumar Deb[1], Pabitra Mitra[2], Madhurima Das[3],

[1]ICAR-Indian Agricultural Statistics Research Institute, New Delhi

[2]IIT-Kharagpur, Kharagpur

[3]ICAR-Indian Agricultural Research Institute, New Delhi

Rice mapping and acreage estimation from regional to global scale level can take several years. In this study, rice area classification and district-level acreage estimation are done and validated with ground-truth data. Rice being a staple food of 3 billion people envisages a huge importance of its mapping and subsequent acreage estimation using remotely sensed satellite data and machine learning. Thus, rice area classification models are developed using Support Vector Machine(SVM), Maximum Entropy Classifier (MEC), Classification and Regression Tree (CART), Naïve Bayes Classifier (NB) and Random Forest (RF)in Google Earth Engine (GEE) and image-based acreage estimation of the mapped rice area was done. Sentinel 2 data and ground truth data were collected from 12 locations in the Cooch Behar District of West Bengal, India from October 2020 to November 2020 during the rice maturity stages The evaluation measures of classification models were done using overall accuracy (OA) and the kappa coefficient. It has been observed that the testing accuracy ranged from 57%(MEC) to 99%(CART) and the kappa coefficient ranged from 0.25(MEC) to 0.97(CART). The developed models accurately estimated rice area with a relative absolute error of 0.45 (MEC) to 0.05 (CART).

Developing Mobile App for Identification of Weeds for Wheat Crop

S.N. Islam, Harsh Sachan, Sudeep Marwaha, Shivadhar Misra, Sudeep Marwaha, Ashraful Haque, Mukesh Kumar, Soumen Pal

ICAR-Indian Agricultural Statistics Research Institute, New Delhi

Information and Communications Technology (ICT) has an important role to play in today's world. Among ICTs, there has been increasing use of mobile phones which has transformed the communication process making a significant impact in many sphere of life including agriculture. Mobile phones have reduced communication and information costs for the rural people. A number of mobile apps have been developed for Indian farmers by different organizations covering various versatile applications for different agriculture issues. Wheat is one of the most important cereal crops after rice. One of the main problems encountered

in wheat crop is the emergence of weeds in the field and our inability to recognize them among native species of plants. The main objective of this research is to provide a mobile application based solution for accurate identification of weed with the help of Artificial Intelligence techniques. This mobile application developed on Android Studio 4.0.0 and the minimum requirement for the app is a device running on API 11. At present, this mobile APP is able to identify five type of weed related to wheat crop that can be extended to more weeds of wheat and other crops as well.

A Transformer-based Computer Vision Model for Recognition of Disease Severity Levels

Md. Ashraful Haque[1], Sudeep Marwaha[1], Alka Arora[1], Chandan Kumar Deb[1] and Karambir Singh Hooda[2]

[1]ICAR-Indian Agricultural Statistics Research Institute, New Delhi
[2]ICAR-National Bureau of Plant Genetic Resources, New Delhi

Generally, transformer models are designed for sequence modelling and transduction tasks in NLP domain. In recent years, tremendous success of transformers has led to their adaptation towards computer vision tasks. In this work, we proposed a transformer-based computer vision model for recognizing the severity levels of diseases using field-collected images. The severity level of diseases is one of the important factors that significantly affect overall crop production. Recognizing the diseases at their early severity level is very crucial for proper crop management. However, traditional lab-based and visual assessment procedures for recognition of disease severity levels become impractical many times due to the requirement of highly trained and experienced personnel. Therefore, we created MDSD, Mayd is leaf blight (MLB) Disease Severity Dataset, containing digital images of MLB disease of maize collected from different agricultural fields. The images of MLB disease were categorized into four classes viz. healthy, low, medium and high severity levels by domain experts. The proposed transformer-based model was trained and tested on the images of MDSD database and reported classification accuracy of more than 94% on the test dataset. The experimental results suggested that the transformer-based networks are quite capable of learning the distinguishable features of the visual images.

Deep Learning-Based Classification Model for Bovine Disease Detection

Madhu, Chandan Kumar Deb, Md. Ashraful Haque, Sudeep Marwaha, SN Islam,
Achal Lama and Mrinmoy Ray

ICAR-Indian Agricultural Statistics Research Institute, New Delhi

Foot and Mouth Disease (FMD), Mastitis, Lumpy Skin Disease (LSD),Non-Clinical Mastitis are the common diseases in bovine. The objectives of this study were to develop classification models for these diseases of bovine using deep learning algorithms with image processing techniques and to identify existing visible illnesses automatically. In this work, we have created an image base of various diseases of Bovine, collected from different farms containing images of four diseaseclassesviz.FMD, Mastitis, LSD, Non-Clinical Mastitis. Therefore, we proposed a deep learning model based on Xceptionnet work for the classification of the images of bovine diseases into respective classes. The proposed model was trained and tested on the collected images of bovine diseases and achieveda classification accuracy of more than 80%. The experiment results showed that the Xception network-based deep learning model is quite capable of disease detection in bovine.

Students Sessions

SP 01:

Data Analytics for Indian Agricultural E-Governance Projects

Ashutosh Prasad Maurya and Pradeep Kumar Suri
Delhi Technological University, Shahbad Daulatpur, Delhi

In India agriculture sector playing an important role from employment and economical angle. For better enhancement of agricultural productivities, supply chain, resource management etc. efforts are being made by the Indian Government through new researches and improvements by Indian Council of Agriculture (ICAR) and through different flagship schemes for farmers directly by Ministry of Agriculture and Farmer’s welfare. These different schemes like Soil Health Card, Agmarknet, Pradhan Mantri Kisan Samman Nidhi (PM-Kisan), Pradhan Mantri Fasal Bima Yojana (PMFBY) etc. are being implemented by use of Information Communication Technology (ICT) under Digital India Programme through e-Governance applications. These collected data from these schemes can be further utilized by data analytics applications to improve service quality, monitoring of activities and planning related to the policy aspects. The possible application of data analytics for agricultural E-Governance projects are in form of dashboards, predictions, anomaly detection, recommendation, process automation, image processing, chat-bots, big data etc. by using different machine learning statistical models. Paper emphasizes about benefits and different possibilities of

analytics in major agricultural E-Governance projects along with technical and data aspects. Integration of information and processing of large integrated dataset using Big Data can be much effective to get better insight for policy decisions. National Informatics Centre is playing a major role for the implementation of E-Governance projects in agriculture through its technical expertise in ICT project implementation. The managerial components like collaborative effort, capacity building, strategic planning, project management etc. can results much more effective application of analytics in agriculture.

A hybrid approach using Genetic Algorithm with K-means in clustering of Indian Mustard genotypes

Hemant Poonia, Ramavtar, B.K. Hooda and Ramniwas
CCS Haryana Agricultural University, Hisar

The study was carried out to check the applicability of hybrid approach using Genetic Algorithm with K-means for clustering of Indian mustard genotypes. The secondary data on growth and yield attributes of 80 Indian Mustard genotypes was used for identification of patterns and best genotypes for plant breeders. A hybrid clustering method was used for getting improved clusters by combining two clustering methods genetic algorithm and k-means. Initially, cluster centers were obtained using genetic algorithm clustering method and then cluster centres were used as input for the k-means procedure. Cluster's quality was measured in terms of total within sum of squares and the sum of squares ratio. It was concluded that improved cluster can be obtained if hybrid clustering method is applied on the subset of selected variables.

On t-packing designs

Sayantani Karmakar[1,] Cini Varghese[1], Seema Jaggi[2] and Mohd Harun[1]
[1]ICAR-Indian Agricultural Statistics Research Institute, New Delhi
[2]Indian Council of Agricultural Research, New Delhi

t-designs constitute a generalized class of 2-designs, more popularly known as balanced incomplete block designs. In t-designs the number of blocks in which any t treatments ($t \geq 2$) occur together a constant number of times. t-designs can find potential application in farming system research where the main concern would be to select the best t-component combination out of available n components ($n > t$) farming systems for a specific agro-ecological zone. However, in order to obtain t-balance, a large number of blocks and replications may be required. Partially balanced t-designs are useful for experimental situations where it is not possible to get t-balanced designs. Partially balanced t-designs form a more restricted subset of t-packing designs. In this study a series of 3-packing designs are

constructed using mutually orthogonal Latin squares. These designs, though not combinatorially balanced in terms of t-tuples, but are highly efficient and have an

increased application potential among experimenters.

Comparison of selection indices using equal weight, standard deviation and heritability weight methods in finger millet through R codes

Radhika L Ghetiya and A.D. Kalola
B.A. College of Agriculture, Anand Agricultural University, Anand

The data for the present study was collected from the experiment conducted in a Randomized Block Design with three replications at the Genetics and Plant Breeding Farm, AAU, Anand during kharif 2020. The experimental material comprised of six different biometrical characters viz. grain yield per plant (g), productive tillers per plant, test weight (g), days to maturity, main ear length (cm) and plant height (cm) of forty genotypes of finger millet. The standard deviation and heritability were worked out between yield and yield attributing characters. The selection index technique was employed to study the crop improvement. Selection indices were calculated for all (sixty-three)possible combinations of six biometrical characters using equal weight (W1), standard deviation (W2) and heritability (W3) as weights and the combination which provided the higher relative efficiency was selected. The expected genetic advance and PRE of different indices were calculated. The results indicated that the selection index I2456(a combination of productive tillers per plant, days to maturity, main ear length and plant height)in standard deviation weight method was found the best. Further the genotypes were ranked based on the selection scores calculated using the b value of the selection indices. The genotypes WN 594, WN 587 and VR 847 were the top three ranking genotypes.

Designs for Breeding Trials using Doubly Nested Partially Balanced Incomplete Block Designs

Nehatai W. Agashe, Cini Varghese, B.N. Mandal and Mohd Harun
ICAR-Indian Agricultural Statistics Research Institute, New Delhi

Breeding techniques are used as tools for the development of commercial hybrids with genetic potential. The information collected on general combining ability effects forms a basis of making correct choice of the best parental lines. Most common and rigorously used breeding techniques are complete/partial diallel, triallel and tetra-allele crosses. Once the sample of crosses are systematically selected, an environmental design is required to assess their performance. Instead, a combined mating-environment (ME) design can be efficiently used for the selection of crosses along with their environmental trial. Balanced incomplete

block designs, partially balanced incomplete block designs, nested balanced incomplete block designs etc. are advantageously being used as ME designs. Doubly nested incomplete block design (DNIBD) is an incomplete block design which includes three systems of blockings, in which first system of blocking is nested within a second system and second system of blocking is nested within a third system of blocking. In this study, sub blocks and sub-sub blocks of doubly nested partially balanced incomplete block design are used to construct ME designs for partial diallel/triallel/tetra-allele crosses.

Study of various Multivariate techniques on Forestry: A case study of Wild Pomegranate (Punica granatum L)

Pawan Kumar1, RK Gupta1, Tara Gupta2 and Sukhdeep Kaur1
Dr. YS Parmar UHF, Nauni, Solan

The study deals with cluster, discriminant, and principal component analysis. The data for the present study were collected. Data on different morphological and seedling characters, namely tree height (m), tree diameter (cm), crown spread E-W (m), crown spread N-S (m), fruit weight (g), leaf length(cm), internodal length(cm), collar diameter (mm), number of branches per plant and leaf petiole (cm) were considered from five different districts comprising of ten seed sources. Two seed sources were considered from each district viz; Narag and Neripul from Sirmour district, Waknaghat, and Sadhupul from Solan district, Sundernagar and Rewalsar from Mandi district, Mohal and Banjar from Kullu district, Basantpur and Sunni from Shimla district. Different morphological and seedling characteristics of ten seed sources were evaluated. Cluster analysis was performed and the seed sources were grouped into three clusters. Discriminant analysis was carried out to categorize the seed sources into high and low yielders. Tree diameter,Tree height, and crown spread E-W are the most important characters that discriminate the two groups. Six seed sources from Mandi, Kullu, and Shimla districts were high-yielders whereas the other four seed sources were low-yielders. Three principal components (PCS) were extracted out of ten which explained 34.675, 23.002, and 11.587 percent of the total variation respectively amounting to 69.29 percent of the total variation.

A comparative study of forecasting models on volatile time series price data

Rabsanjani Pramanik1*, Md. Wasi Alam2*
1ICAR- Indian Agricultural Statistics Research Institute, New Delhi
2ICAR- Indian Agricultural Statistics Research Institute, New Delhi
*Corresponding author Email: rabsan12@gmail.com

Efficient and reliable forecasting techniques for commodities with volatile price series are indispensable in agriculture dependent country like India. In this context, choosing a universally accepted model for forecasting precisely the price series of commodities like onion is one of the most challenging tasks because of the existence of seasonality, non-linearity and complexity in the data simultaneously. Time series models like GARCH, machine learning techniques like TDNN, SVM and deep learning models like LSTM, Stacked LSTM and Bi-LSTM have been extensively studied in this research work to judge their performance on volatile weekly price series of onion for two different markets in India. The models were tuned with the training dataset to forecast the values for the next twelve horizons and eventually the forecasted values have been compared with the testing dataset. As seen, deep learning models outperform the machine learning and time series model in dealing with the two datasets.

Bootstrap prediction intervals in time series for neural network models

Samir Barman*, Ramasubramanian V., Mrinmoy Ray, K.N. Singh and Anshu Bhardwaj

ICAR-Indian Agricultural Statistics Research Institute, New Delhi

Time Series(TS) modelling is an important area in the domain of statistics, as it enables us to comprehend the dynamics underlying a particular phenomenon. One may obtain TS forecasts as either point or interval forecasts. The interval forecast contributes to a comprehensive assessment of future uncertainty for improved risk management choices, the planning of diverse strategies for the range of potential outcomes, and the exploration of scenarios based on various assumptions. There are few methods described in the literature for constructing prediction intervals (PIs) for artificial neural network (ANN) models utilising the Bootstrap method. The existing bootstrap approaches take into account the variability resulting from bootstrap estimation of parameters and but not that resulting from model estimation of parameters. To also account for the latter, predictive root concept earlier used by researchers for both linear and non-linear auto regression models has been extended to ANN models for constructing PIs. Two bootstrap approaches (with and without rescaling) for constructing PIs in ANN models for non-linear T Shave been proposed. The performance of the proposed methods have also been evaluated by comparing it with the existing methods using both simulated and real datasets. The proposed methods can be considered as a viable alternative for computing PIs in time series.

Spillover Effects of Covid-19 Induced Lockdown on Onion Prices in India

Sandip Garai*and Ranjit Kumar Paul
ICAR-Indian Agricultural Statistics Research Institute, Pusa, New Delhi

In a normal situation onion price varies in a highly unprecedented way in India. So, it is worth noticing the effect of an uncertain situation on the onion price. In this article an efficient Artificial Intelligence (AI) tool, i.e., Support Vector Regression (SVR) has been used to predict the price fluctuation of onion over the lockdown period, unlock condition and the period including the pre-pandemic situation. Results obtained are compared with prediction of traditional Multiple Linear Regression (MLR)model. Several metrices such as R^2R^2, Root Mean Squared Error (RMSE), Mean Absolute Deviation (MAD), and Relative Mean Absolute Percentage Error (RMAPE) have been used for this purpose. The result of Machine Learning (ML) algorithm indicates that in the nationwide lockdown condition pandemic indicator variables are having more than 70% influence on the onion price variability. The effect is reduced to near about 60% in unlock condition and if considering the whole year data this effect is near about 45%. The results also indicate that ML algorithm is more efficient to capture the variability than the traditional model.

Agenetic algorithm optimized deep neural network for wheat yield prediction using spectral indices

M. Lokeshwari
ICAR-IASRI, New Delhi

Crop yield prediction is of great interest for market behavior, to increase global food security, food trade, and to drive governmental policies. Early estimation of yield at field or plot scale also contributes to high-throughput plant phenotyping and precision agriculture. The objective of this study is to evaluate the power of spectral, thermal and structural information to estimate wheat grain yield within the framework of a Deep Neural Network (DNN).Existing approaches had several drawbacks, including low accuracy, increased computational complexity, higher feature dimensionality, poor reliability, and increased time consumption due to poor hyperparameter optimization. In this paper, we proposed a Genetic Algorithm (GA)-based DNN optimized architecture for the early estimation of crop yield. The GA algorithm is established to optimize the hyperparameters in DNN architecture such as the number of hidden layers, the number of neurons in the hidden layer, the learning rate value, and the batch sizes. Multimodal information, such as canopy spectral, structural, and thermal features, were combined to predict crop grain yield using Least Absolute Shrinkage and Selection Operator(LASSO), Random Forest Regression (RFR), Support Vector Regression (SVR)and Genetic Algorithm optimized Deep Neural Network (GA-DNN).The results can be summarized in three messages: (1) Hyperparameter tuning improves the yield prediction accuracy and is more adaptable to spatial variations; (2) DNN-based models improve yield prediction accuracy: the highest accuracy was obtained by GA-DNN with an R2 of 0.84 and root mean square error (RMSE) of 98.57; (3) DNN-based models were less prone to saturation effects and exhibited more adaptive performance in predicting grain yields across the wheat germplasm in our study. Furthermore, the result also revealed that spectral features had a greater effect on crop yield than canopy thermal features.

SP 02: Students Sessions

Transformer based neural network for agricultural price forecasting

G. Avinash, Ramasubramanian V., Harish Nayak G.H. and Mrinmoy Ray

ICAR-IASRI, New Delhi

Computer vision, self-driving vehicles, product recommendations, behavior analysis, natural language processing and medicine are just a few of the domains that have been greatly improved by deep learning. The financial sector is no surprise where the use of deep learning has produced one of the most lucrative applications. This research proposes a novel Deep Learning method that uses Transformer neural networks for agricultural price forecasting. They have the ability to identify the dynamics and complex patterns from irregularities in financial time series. In this study, Transformer architecture consisting of input layer with positional embeddings, self-attention with layer normalization has been used and the resultant output has been passed to the deep neural networks. Multi-head attention mechanisms have been utilized within the transformer to learn non-linear patterns and dynamics with high volatility present in the time series data. Transformer model was utilized to analyze the Soya price series data and the performance of the model was compared with other deep neural networks like Gated Recurrent Unit (GRU), Long-Short Term Memory (LSTM), Bidirectional LSTM (Bi-LSTM). The models were evaluated using RMSE and MAPE and the results revealed that the performance of Transformer neural network was superior compared to other models.

STL-LSTM Hybrid Model for Forecasting Seasonal Agricultural Price Series

Ronit Jaiswal1 and Girish K. Jha2

ICAR-IARI, New Delhi

In this study, we develop an ensemble hybrid model called STL-LSTM for a non-stationary, nonlinear and seasonal agricultural price series by integrating a decomposition technique viz. seasonal trend decomposition procedure based on loess (STL) with an efficient recurrent neural network-based forecasting technique, i.e. long short-term memory (LSTM). The STL technique decompose the original price series into its seasonal, trend and remainder components whereas the LSTM network is constructed for each component to forecast them individually. Finally, the prediction results of all components are aggregated to formulate an ensemble output for the original agricultural price series. The hybrid model captures the temporal patterns of a complex time series effectively through analysis of the simple decomposed components. The study further compared the price forecasting ability of the developed STL-LSTM model with the other potential models using monthly price series of potato for two major markets of India. The empirical results demonstrated the superiority of the developed hybrid model over the other models in terms of root mean square error (RMSE) and mean absolute percentage error (MAPE). Moreover, the evaluation of the accuracy of the forecasts using the Diebold-Mariano test confirms that the STL-LSTM model has a clear advantage over the other models.

Forecasting of Groundnut Yield using Meteorological Variables in Junagadh, Gujarat

P.C. Bhimani and H. V. Gundaniya

B.A. College of Agriculture, Anand Agricultural University, Anand

Agriculture is one of the most significant economic sectors in India, which is strongly dependent on climatic conditions. Groundnut is one of the major oilseed crops of Gujarat and Junagadh has highest production of groundnut in Gujarat. For effective, forward-looking, and current planning, especially in agriculture, which is fraught with uncertainty, reliable and timely forecasts are essential. Therefore effective yield forecast of such important oilseed crop is necessary for future planning and policy making. In the present investigation, the time series data of groundnut yield and weather parameters of 29 years Junagadh (1991-92 to 2019-20) were used. The week wise weather indices were generated using correlation between de-trend yield and weekly weather variables. Multiple Linear Regression (MLR) and Discriminant Function Analysis were used to develop yield forecasting model using weather indices before three week and one week before harvest, respectively. These models were compared using Coefficient of multiple determination (Adj. R2), Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). The study found that model-1A, developed using MLR technique have high value of adj. R2 (78.8%) and low RMSE value (671.72). Multiple linear regression (MLR) was found to be more accurate than the discriminant function analysis approach.

A Glance to Bioinformatics: Tools to Mitigate Covid-19

Shilpa Sharma, RK Gupta, Ashu Chandel, Anju Sharma

College of Forestry, Dr YSP UHF, Nauni, Solan

At the end of 2019, the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic, also known as coronavirus disease 2019 (COVID-19), was made public. The severity of the pandemic and its global spread prompted a previously unheard-of effort from the scientific community, and a great deal of new research was conducted, particularly by researchers in the fields of medicine, biology, public health, bioinformatics, and computer science. The primary method for locating the SARS-CoV-2 genomes and obtaining the

fundamental information about the virus is next-generation sequencing. Key tools for processing such data and enhancing understanding of SARS-CoV-2 include biological and host-virus interaction databases. This research discusses the role of first-strand bioinformatics in reducing the negative effects of Covid-19. In Computational techniques to battle COVID-19: practical tools to speed coronavirus and SARS-CoV-2 research, In order to give crucial tools for the identification, comprehension, and management of COVID-19, Franziska Hufsky et al. propose bioinformatics tools that have been specifically created for SARS-CoV-2. The tools that were evaluated include SARS-CoV-2 detection, sequencing data processing, tracking and containment of the COVID-19 pandemic, coronavirus evolution research, identification of prospective medication targets, and associated therapeutic approaches.

Variance Components Estimation and BLUP in Breeding Programmes under P-rep block Designs

Vinaykumar L.N.1, Cini Varghese2, Seema Jaggi3 and Mohd Harun2
1ICAR-Indian Agricultural Research Institute, New Delhi
2ICAR-Indian Agricultural Statistics Research Institute, New Delhi
3ICAR, New Delhi

One of the most important aspects in the Early Generation Breeding Trials (EGBTs) of any breeding programme is the effective use of available resources. EGBTs typically involve multiple environments such that in each environment, a large number of lines ($v > v > 1000$) are to be tested. Use of partially replicated (p-rep) block designs, where a proportion of the test lines are replicated in each environment, can be used efficiently in such multi-environmental trials. Since, only a sample of inbred lines from a possibly large hypothetical population is considered, the use of random effects model would give more reliable results. Variance components estimates are needed for better prediction. n this study, the problem of estimation of variance components and prediction of the unobserved line effects along with general mean effects has been considered under a p-rep block design set up.

Performance Evaluation of Machine Learning Algorithms for Wheat Crop Production Utilizing Soil Parameters

Afshan Tabassum1,M. Iqbal Jeelani1, Manish Sharma1,Khalid Ul Islam Rather1, Imran Rashid and Mansha Gul

1 SKUAST-Jammu

Crop production is significantly influenced by the soil and making advance plans in this context can help farmers and soil analysts to increase crop production by using the results of soil parameters. In this framework secondary data of diverse soil parameters and production with respect to Wheat crop from various districts of Jammu region were collected from different experimental setups in collaboration with the Soil Testing Lab, Talab Tillo Jammu. Four supervised machine learning algorithms including K-Nearest Neighbour (KNN), Support Vector Regressor (SVR), Random Forest and Linear Discriminant Analysis (LDA) were evaluated based on the selection criteria's like Accuracy, Kappa, Prevalence, Sensitivity, Specificity and Detection rate for prediction and validation of wheat production utilizing 80:20 cross validation approach with the help of various libraries of R studio version 4.2.1.Based on the results of selection criteria, Random Forest machine learning algorithm out performed rest of the algorithms used in the current study.

Statistical Analysis of Rainfall Patterns n Chhatarpur District of Madhya Pradesh

Dagam Sindhu[1] and R.B.Singh[2]
Jawaharlal Nehru Krishi Vishwa Vidyalaya, Jabalpur

Statistical analysis of rainfall has been of great concern during the past century as well as today because of the consideration has been given to global climate change by the scientific community. For this purpose, the monthly rainfall and rainy days for 25 years i.e., 1987-2011 of Chhatarpur district has been collected for trend analysis from Dept. of Agro. meteorology, JNKVV, Jabalpur, M.P. The present study attempted to know the trend of rainfall and rainy days. The change point analysis has been analysed for observing any change. Trends of rainfall parameters were analysed using Mann-Kendall test, Sen's slope test, Spearman's rho test, regression and change point analysis. The average annual rainfall for the district during the study period was found to be 923.48mm with a coefficient of variation of 2099.1 percent. This study revealed that significant decreasing trend in annual, monsoon and September rainfall have been observed at Chhatarpur, whereas May month shows increasing trend of rainfall. Though, rainfall in the month of May has increased significantly at 5% level of significance but there is no significantly trend in rainy days. From the period 1987-2011, pettitt change point analysis revealed that there is no significant change in rainfall and rainy days.

SP 03: Students Sessions

Processing 3D Point Clouds using Open3D library for High-Throughput Plant Phenotyping

Preety Dagar1, Alka Arora1, Sudeep1, Rajni Jain2, Mrinmoy Ray1, Vishwanathan Chinnusamy3 Sudhir Kumar3

ICAR-Indian Agricultural Statistics Research Institute, New Delhi

2ICAR-National Institute of Agricultural Economics and Policy Research, New Delhi

3ICAR-Agricultural Research Institute, New Delhi

In recent years, measuring three-dimensional (3D) surface information has gained a great interest in plant phenotyping because it can represent the nature of plant architecture better than conventional 2D images. In this paper, we have used Open3D, a three-dimensional (3D) open source Python library to process 3D information. High resolution RGB multi-view imagery collected using a mobile camera was used to reconstruct 3D model of canopy of a plant. A 3D point cloud of the sample was constructed and analyzed. Later sophisticated 3D operations were performed on these images including 3D reconstruction, voxel down sampling, estimation of point normal and generation of depth maps etc. The information generated can enhance the evaluation of crop traits and provide accurate statistics for assessment of their growth parameters. The goal of this paper is to use Open3D library to process 3D point clouds of a plant canopy so that further it can be used to extract plant phenotyping features.

Vocational education and its employability perspectives: A learning ecosystem analysis of fisheries vocational courses in Kerala

Reshma Gills*, C Ramachandran, V P Vipinkumar, Eldho Varghese, Pooja Krishna and T V Ambrose
ICAR-Central Marine Fisheries Research Institute, Kochi

Human resource development is critical for sustaining, diversifying, and realizing the potential of fisheries, like any other sector in agriculture. The fisheries sector (both capture and culture) is projected as a potential area that offers numerous employment-oriented avenues due to the fast pace of transformation. National Education Policy, 2020 realized that vocational education in this sector is the basic foundation for the development of skilled human resources as it plays a major role in the entrepreneurial molding of youth, especially in coastal areas, to fill the expected gap of expertise. Though vocational fisheries education is projected as a potential employment supportive initiative, retrospective information regarding the scale of employability and validated research data on the status of its present attainments is very sparse. This paper attempts to analyze the learning ecosystems of fisheries vocational courses in Kerala in the context of their perceived employability. In the present study, a total of 964 students and 36 teachers from 18 vocational higher secondary schools from three different streams of vocational fisheries courses [Aquaculture (Aqua), Marine Fisheries & Seafood Processing (MFSP), and Marine Technology (MT)] participated. The study results

highlighted the perceived employability status, learning motivators, and constraints in the current fisheries vocational education system.

A Regression Type Estimator In Stratified Adaptive Cluster Sampling With one and two Auxiliary Variables

Mhalo Ovung and Phrangstone Khongji

North-Eastern Hill University, Shillong

In this paper, we construct an estimator of the population mean employing stratified adaptive cluster sampling using auxiliary variable information. As the use of an auxiliary variable enhances the precision of the estimator, in this present work we constructed a regression type estimator. The estimator is found to be biased and accordingly the mean square error was obtained. The efficiency of the estimator is compared with some existing estimator both analytically and numerically. In case of the former the efficiency is conditional on the correlation coefficient. In case of the later, through simulated data, the proposed estimator is more efficient compared to the Ratio estimator in stratified adaptive cluster sampling based on Hansen-Hurwitz given in Chutiman [1] and usual stratified adaptive cluster sampling given in Thompson [2].

In-silico characterization and molecular modeling of UDP Glucose Pyrophosphorylase in Rice

Mamatha Y S1,Sudhir Kumar2, Neeru Redhu2 and Dwijesh Chandra Mishra1

ICAR-Indian Agricultural Statistics Research Institute
Chaudhary Charan Singh Haryana Agricultural University

The enzyme UGPase functions as a catalyst in the reaction of Mg+2-UTP + Glu-1-P PPi + UDP-Glu, which results in the synthesis of sucrose. Depending on the tissue's metabolic state, it can catalyse both forward and backward reactions. The catalytic domain of UGPase has a nucleotide binding loop in the middle, along with the N- and C-terminal domains. In the vast majority of photosynthetic plants and microorganisms, UGPase is functionally active in its monomer form. In order to compare the structural modelling of UGPase, models from Modeller, Swiss-Model, and I-TASSER were used. To find the best stable model, all predicted models were further examined, analysed, and improved using the QMEAN, WHATIF, and SAVES servers. Protein-protein docking was used to construct homodimers in order to explore the interaction between monomer units. H Bond, NACCESS, and CastP server were used to predict hydrogen bonding between monomer units, solvent accessible area, and ligand binding/active sites, respectively. In comparison to the monomer, the solvent accessible region in the dimer was around 100–200 smaller. This study showed that a

reduction in solvent accessible area and a subsequent smaller active site resulted in a loss of activity due to dimerization in UGPase.

Genome wide identification and characterization of long non-coding RNAs and circular RNAs in Black pepper (Piper nigrum L.) and studying their role as competitive endogenous RNAs

Baibhav, Vinayaka and M.A. Iquebal

ICAR-Indian Agricultural Statistics Research Institute, New Delhi

Long non-coding RNAs (lncRNAs) and circular RNAs (circRNAs) are two classes of non-coding RNAs (ncRNAs) present predominantly in plant cells and have various gene regulatory functions at pre and post-transcriptional levels. Previously deemed as “junk” these ncRNAs have now been reported to be an important player in gene expression regulation, especially in stress conditions in many plant species. Black pepper despite being one of the most economically important spice crop is lacking such ncRNAs related studies. We analyzed 53 raw sample data of black pepper to identify lncRNAs in 6 tissues viz. flower, fruit, leaf, panicle, root and stem identifying a total of 8285 lncRNAs and characterized in flower, fruit, leaf, panicle, root and stem tissues. Further downstream analysis inferred that the lncRNAs could regulate 157 black pepper genes or gene products via miRNA-lncRNA-mRNA network interactions thus working as ceRNAs. The interactions can be various mechanisms like miRNA-mediated gene silencing or lncRNAs acting as endogenous target mimics (eTMs) of the miRNAs. 5 lncRNAs were also identified to be potential precursors of 20 miRNAs after being acted upon by endonucleases like Drosha and/or Dicer. Tissue-wise transcriptome analysis revealed 1045 circRNAs in flower, fruit, leaf, panicle, root and stem tissues respectively. Further miRNA-circRNA-mRNA network analysis showed 171 circRNAs combining with 42 miRNAs and competing for the binding sites on 182 mRNAs in different black pepper tissues. The findings we offer through the study can help us to get a better understanding of the black pepper’s yield regulation and responses to stress and subsequently helping the scientific community to further breed more vigorous black pepper varieties.

Machine learning in the estimation of CRISPR-Cas9 cleavage sites

Jutan Das and Sanjeev Kumar

ICAR-IASRI, New Delhi

CRISPR-Cas9 system is one of the most used genome editing techniques in the recent time. In spite of its high potentiality to modify the specific target genes and region of the genome which are complementary of the designed guide RNA (or sgRNA), still it suffers from the off-

target effect. Here, in this study, an attempt has been made to develop models based on three machine learning based techniques (i.e. Artificial Neural Network, Support Vector Machine and Random Forest) for estimation of the CRISPR-Cas9 cleavage sites to be cleaved by a given sgRNA. All these machine learning based models were exclusively developed on the plant dataset. The models were trained on the 70 percent of the collected on-target and off-target dataset of different plant species. Whereas the performance of the model were evaluated on remaining 30 percent of collected data based on following statistics; specificity, sensitivity, accuracy, precision F1 score, F2 score and AUC. All together eleven models were trained based above machine learning techniques. Relative evaluation of these developed models reveals that model based on random forest technique shows better performance. Random Forest model accuracy is 96.27% and area under ROC curve (AUC) was found to be 99.21%. Total six models based on ANN technique (ANN1-Logistic, ANN1-Tanh, ANN1-ReLU, ANN2-Logistic, ANN2-Tanh, and ANN-ReLU) and four SVM models (SVM-Linear, SVM-Polynomial, SVM-Gaussian and SVM-Sigmoid) were trained. The performance of ANN1-ReLU and SVM-Linear model were found to be better among ANN and SVM based models respectively but overall performance of Random Forest is better.

Quantum Chemical Calculation And In Silico Molecular Modelling Studies On Some Multi- Targeting Anti-Inflammatory Inhibitors

Ragini Kushwaha, Sandeep Pokharia, Rajeev Mishra

ICAR-IASRI, New Delhi

Computer-aided Molecular Design helps to discover the quantitative structure activity relationship (QSAR) of a molecule. One of the approaches to create QSAR is to compute quantum chemical descriptors and for that Density Functional Theory (DFT) has been came out. Apart from this In Silico molecular docking studies have also been carried out to estimate the preferred orientation and binding affinity between two molecules (Protein-ligand). Further In Silico prediction of ADMET properties to formulate the pharmacokinetics and pharmacodynamics of molecules. We know that Inflammation pathways may lead to the development of diseases like Rheumatoid Arthritis. Within many inflammation cascades there are often pivotal molecular targets such as Interleukin(IL)-17, 18, Cytokines TNFα, IL-6, IL-1β, Cyclooxygenase, 5-lipoxygenase and MAPKs that, when neutralized or antagonized, block inflammation. In this study four set of natural compounds (Oleocanthal, Viridicatin, Liquiritin, Nobiletin) have been taken which have anti-inflammatory activity. By using the Gaussian03W/09 software, all the quantum chemical calculations have been evaluated based on DFT B3LYP/6-311G(d,p) level. For molecular docking and ADMET studies AutoDock tools-1.5.6 and Molinspirationhave been used respectively. A comparative analysis of calculated parameters have been carried out. Thus more effected drug can be developed using combined quantum-chemical and In Silico drug designing approaches.”

Weather based Modelling of Rice Yield using Statistical & Machine Learning Models

Akhilesh Kumar Gupta, Kader Ali Sarkar, Digvijay Singh Dhakre, Debasis Bhattacharya

Visva-Bharati, Sriniketan

Crop yield modelling and forecasting has been the crucial step for agricultural as well as economic policy decision in India for a long time. It is critical requirement for better planning and policy decisions relating to storage, procurement, distribution, pricing, marketing, export-import etc.The present study emphasized on understanding the relationship between crop yield and different weather variableswith an aimto develop a reliable crop yield prediction system for Rice (Oryza sativa L.) yield in West Bengal. The effect of weather variables on rice yield was studied by the means of weather indices whichare assimilated representation of weekly weather effects on rice yield. Various statistical and neural network models were developed based on the linearity and non-linearity of the data. The best model was selected on the basis of lowest RMSE, MAE and MAPE. The resultof statistical models shown that the effects of rainfall, minimum temperature and time variable tare significant and there is presence of both linear and non-linear pattern in the data. The accuracy of neural network models was better than statistical models and the study revealed that the hybrid models with 3 and 4 hidden nodes were best fitted model in Burdwan and Birbhum districts respectively. The best fit models can be used to obtain a reliable forecast of rice yield at 6-8 weeks before harvest for various policy decisions.

An Analysis of Disparity among the Potao Producing Districts of West Bengal using Markov Chain Model

Mamata and D. Bhattacharya
PSB, Visva Bharati university, Santiniketan

This paper concentrating in studying the disparity among the districts growing potato during 1991-92 to 2020-21 for the districts of West Bengal. Seventeen districts have been considered for this study. This paper analyzing thirty years of potato production data. On the basis of the potato production data districts have been classified into three states, namely, highly developed, developed and under developed. Next observing the frequency of change in the states the transition probability matrix and initial probability vectors are obtained. The steady state probability and expected return time to a particular state are also obtained. Here present stationary probabilities of different districts under study have been used to predict the future movement of the districts from one classification to other in terms of potato production.

Resampling Method of Variance Estimation for Two Stage Sampling under Dual Frame Surveys

Moumita Baishya, Tauqueer Ahmad and Ankur Biswas
ICAR-Indian Agricultural Statistics Research Institute, New Delhi

Multiple Frame (MF) Survey is an approach to combine two or more sampling frames instead of a traditional single frame of units from the target population. The main purpose of using MF surveys is to reduce cost while maintaining estimation efficiency almost at par with single frame surveys. In many cases, a frame that covers the entire population is very expensive to sample from.So, an alternate frame may be available that does not cover the entire population but it is cheaper to sample from.For multi-stage designs, the theory of multiple frame surveys (Hartley, 1962, 1974) becomes somewhat complicated as the alternatives for multiple frame approach multiply, e.g. in two stage sampling there may be multiple frames at the first stage and single frame at the second stage or vice-versa. There may be situations where more than one frames are available at each stage of selection.Saxena et al. (1984) extended Hartley’s estimator for two stage sampling when domain size is known.Hartley (1962) noted thatthe unbiased variance estimation for estimation of various population parameters is difficult and complicated for multiple frame surveys as compared to sample from single frame.Unbiased variance estimatoris very tedious to obtain for estimator using multiple frame as the estimator is non-linear and complex in nature. Therefore, alternatively we can use resampling methods of variance estimation e.g. resampling technique developed by Ahmad (1997) in case of multiple frame surveys under two stage sampling design for known domain size case.

In order to avoid the complications, Kumar et al.(2021) developed different rescaled unbiased variance estimation procedures for estimating varianceunbiasedly for simple random sampling under dual frame surveys namely, (i) Stratified Rescaling Bootstrap with Known Domain size (StRBKD) and (ii) Post-stratified Rescaling Bootstrap with Known Domain size (PstRBKD). In this study, we have developed rescaled unbiased(asymptotically) variance estimation procedure for variance estimation of population total unbiasedly for two stage sampling under Multiple Frame Surveys. Further, statistical properties of the proposed bootstrap variance estimation procedure have been examined through a simulation study. Further, to evaluate the performance of the developed procedure with standard existing procedures without considering any rescaling factors,a simulation study has been carried out and Percentage Relative Bias (RB) and Relative Stability (RS) have been calculated to compare the performance of the proposed bootstrap variance estimation technique.

Weather Intervention based Sugarcane Yield Forecasting Models for Enhancing Farmers’ Income

H.V.Gundaniya1 and Y.A. Garde2
1B.A.College of Agriculture, Anand Agricultural University, Anand
2N.M. College of Agriculture, Navsari Agricultural University, Navsari

Agriculture is one of the most important sector of an Indian Economy. Sugarcane is one of the most important commercial crops grown in India. Pre-harvest forecasts have significant value in agricultural planning and play important role in doubling the farmers' income. Therefore proper forecast of such important commercial crop is necessary for future planning, policy making and sustainable production. In the present investigation, relationship between yearly cane yield and weekly weather parameters were studied by using Karl-Pearson's correlation coefficient approach. The study revealed that all the weather variables were significantly correlated with cane yield in different weeks of cropping season. The sugarcane yield forecasting models were developed using 26 years of cane yield and corresponding weather data (1991-92 to 2016-17). The statistical tools of multiple linear regression (MLR) and discriminant function analysis were used for model development. The study showed that Model-1h, developed through MLR technique have high R2value (92.6%) and low value of RMSE (6.42) as compared to remaining models. Therefore, study concluded that multiple linear regression(MLR) was more reliable as compared to discriminant function analysis approach and which provide yield forecast well in advance of actual harvesting of the crop.

Identification of bacteriophage from the metagenomic data of Ganga and Yamuna rivers

Soutrik Mukherjee
ICAR-IASRI, New Delhi

Microbes are very important in every aspect of not only human but also all the life-forms in earth. Each and every system in the biosphere is induced by the almost infinite ability of microbes to transform the world around them. Identification of bacteriophage from various regions of Ganga and Yamuna rivers was indeed a very important task to know the abundance of different species of bacteriophages. As bacteriophages play a very important role in riverine system by checking the growth of bacteria, it was very important to understand abundance of bacteriophages. Further, very few works have been done on the annotation of bacteriophages identified from the Ganga and Yamuna rivers. Sediment samples from various regions of Ganga and Yamuna River like Balkeshwar-Shivpuri-Agra, Koteswar-Ganga, Rasulabad-Ganga, Sahi-Dabad-Ganga, Taj-Gung-Yamuna, Triveni-Sangam-Ganga, Yamuna- Expressway-Agra, Bagwan-Ganga area by ICAR-

Central Inland Fisheries Research Instituteunder CABIN project. Two approaches were followed for the identification of bacteriophages, one is identification of bacteriophages by binning of the metagenomic contigs data with Metabat2 tool and then distinguishing bacteriophage sequences by a machine learning based tool MARVEL. The other approach was alignment-based approach by BLASTN with the query as the contigs of the metagenomics samples and database made from the bacteriophage sequences downloaded from NCBI. With MARVEL tool from the 9 datasets, two bins of Balkeshwar-Ganga contigs data shows the result of having bacteriophage sequence.Using the bioinformatics software program Blast2GO, unique sequence data was automatically and quickly functionally annotated (genes, proteins).Table describes the quantification of the bacteriophage species from the samples was generated. Aeribacillus phage AP45, complete genome phage was the most abundant phage in all 9 sites of Ganga and Yamuna rivers and gene ontology pie chart describes the biological process, cellular component and molecular function.

A two-stage GWAS analysis method using machine learning approaches for detection of non-marginal epistatic interactions in crops

Tanwy Dasmandal, Dipro Sinha, Sunil Archak and Dwijesh Chandra Mishra
ICAR-IARI, New Delhi
ICAR-IARI, New Delhi
ICAR-NBPGR, New Delhi
ICAR-IASRI, New Delhi

Dramatic rise in genomic data across organisms has made Genome Wide Association Studies (GWAS) a very reliable approach to uncovering associations between genetic variants and complex traits among organisms. The traditional ‘one locus at a time’ approach can only explain a little part of the genetic variance associated with the phenotype. The remaining part can be explained by epistasis which plays enormous role in the understanding of complex crop-trait genetics. Several statistical as well as machine learning approaches have been developed in the recent past to detect epistatic interactions. However, the bottlenecks of the existing approaches are their inability to accurately detect non-marginal epistatic interactions and their limitation to detect only pairwise interactions. To deal with these issues we have made an effort to develop a two-stage GWAS analysis method for accurately detecting non-marginal epistatic interactions in crops. In our work we have applied machine learning approach to filter out the most important SNPs associated with phenotype of interest which were then investigated for interactions using information gain strategy. The filtering strategy has facilitated the detection of multi-locus interactions by reducing the computational burden. Our method has been found to successfully detect non-marginal epistatic interactions for all the simulated models.

An Integrated Feature Selection Framework for Accurate Prediction of 6mA Sites in Crops

Dipro Sinha, Tanwy Dasmandal, Dwijesh Chandra Mishra and Sunil Archak
ICAR-IARI, New Delhi
ICAR-IASRI,New Delhi
ICAR-NBPGR, New Delhi

N6-methyladenine (6mA) is one of the most common epigenetic modifications involved in a diverse range of biological functions in all three kingdoms of life. Despite being more accurate in identifying epigenetic changes, in-vitro methods take a lot of time and resources. These barriers have been surmounted with the use of in-silico techniques based on artificial intelligence. To achieve more accurate results, an effort has been made to build a machine learning framework with an integrated feature selection module on two model plants viz. rice and Arabidopsis. In our work, Di-Nucleotide Frequencies, GC content, and Average Mutual Information Profile were first chosen as three distinct feature sets. A combination of two feature selection strategies were then adopted on these feature sets to obtain the final features for 6mA site prediction. Three machine learning models have been tested using different feature sets viz. the three feature sets separately, combinations of the three feature sets and the final feature set obtained using the feature selection module. The results showed that all the three machine learning models performed better in terms of accuracy, sensitivity, specificity, Mathew’s Correlation Coefficient and Area Under Curve using the integrated feature selection module as compared to other feature sets.

Stock Reduction Analysis for Assessment of Data-limited stocks: A Case Study of Marine Fishery in Andaman Nicobar Islands

Judith Das1*, Nesnas Nazir2, SreepriyaV.1, Eldho Varghese1, J. Jayasankar1 and Mini K. G.1
1ICAR-Central Marine Fisheries Research Institute, Kochi
2Dept. of Fisheries, Port Blair, UT of Andaman & Nicobar

Depletion-based stock reduction analysis (DB-SRA) is an approach used to study data-limited fisheries and usually provides robust results with minimum input requirements. The study utilizes catch along with some life history parameters as the input data and provides the historic carrying capacity (K) and the overfishing limits (OFL), which can be used to identify the status of the stock and its maximum sustainable yield, which forms the key parameters for recommending the catch limit for the sustainable management of the fishery. In this paper, we have attempted to explore the SRA approach for assessing a few major stocks in the Andaman and Nicobar Islands. A comparative analysis was also made with other popular data-limited stock assessment methods.

Post Stratified Separate Ratio Exponential Estimator In Two Stage Sampling Design

M.Younis Shah, S.E.H. Rizvi, M. Sharma And M.Iqbal Jeelani

Division of Statistics and Computer Science, Faculty of Basic Sciences, Main Campus
SKUAST-J, Jammu, J&K,

Post-stratification in two-stage sampling using auxiliary data at the second stage has been tried. Units are chosen using simple random sampling without replacement at both stages. In this research, we propose a post stratified separate ratio exponential estimator in two stage sampling design. Expressions for the proposed estimator's bias and mean square error (MSE) up to the first degree of approximation are developed. Additionally, the optimal constant value that minimizes the MSE of the suggested estimator is determined. Mehrotra and Kumar investigated the issue of post stratification in two-stage sampling with post stratification on the basis of ultimate-stage units. Empirical research has been done to show that the suggested estimator is more efficient than other estimators that have been taken into consideration. The performance of the modified estimator is assessed using the proposed MSE and Percentage Relative Efficiency (PRE).

SP 05: Students Session

Forecasting of Pigeon pea yield in Bihar through Auto-Regressive Integrated Moving Average Model (ARIMA)

Naveen and Mahesh Kumar
Dr. Rajendra Prasad Central Agricultural University, Pusa, Samastipur

This research paper entitled “Forecasting of Pigeon pea yield in Bihar through Auto-Regressive Integrated Moving Average Model (ARIMA)” is based on the secondary data. Data was collected for the years 1980 to 2021 from the websites and Department of Economics and Statistics and India Agri. Stat. The different ARIMA models are judged on the basis of Autocorrelation Function (ACF) and Partial autocorrelation Function (PACF) at various lags. The ARIMA models are fitted to the original time series data as well as the first difference data to check the stationarity. The possible ARIMA model are identified on the basis of significant coefficient of autoregressive and moving average components. Data for the year 1980 to 2019 were used for model building and rest two years data i.e. 2020 and 2021 were used for validation of yield forecast model of Pigeon pea .

It was found that the ARIMA (1,1,1) model is best fitted for forecasting of Pigeon pea yield in Bihar among all the other models namely ARIMA (0,0,1), ARIMA (1,0,1), ARIMA (0,1,1),

ARIMA (1,1,1), ARIMA (0,1,2), ARIMA (2,0,0) and ARIMA (2,0,1). The parameters of all these models were computed and tested for their significance. Various statistics were also computed for selecting the adequacy, invertibility, stationarity, Diagnostic checks, % forecast error and parsimonious model with t-test and chi-square test. The low values of MAPE, MAE, RMSE, % forecast error and BIC are good inferences for forecasting of Pigeon pea yield in Bihar. Five years ahead forecasting from 2020 to 2024 for pigeon pea yield were calculated using ARIMA models. The five year ahead yield forecasting of pigeon pea in Bihar for the year 2020,2021,2022 ,2023 and 2024 are 1723 kg/ha, 1680kg/ha, 1671kg/ha,1693 and 1707 kg/ha respectively with forecast error for the year 2020 and 2021 are 2.23% and 4.62% respectively.

The ARIMA model for forecasting yield of Pigeon pea in Bihar is as below:

$Zt – Zt\text{-}1 = 16.019 + 0.485\ (zt\text{-}1 - zt\text{-}2) – 0.995\ (at\text{-}1 - at\text{-}2) + at$ (for Bihar)

Yield forecasting of Chickpea in Bilaspur district (Chhattisgarh) using nonlinear models

Kalyan Singh Paikra and Mahesh Kumar
Dr. Rajendra Prasad Central Agricultural University, Pusa, Samastipur

This study is done with objective of forecasting chickpea yield for Bilaspur district (Chhattisgarh) using nonlinear growth models. For conducting this study, secondary time series data is obtained from official sites of Directorate of Economics and Statistics, Chhattisgarh and ICRISAT, Hyderabad from time period of 1995-96 to 2019- 20. For achieving objective, data from 1995-96 to 2018-19 are analysed while for validation, data from year 2019-20 is taken. There different nonlinear growth models namely Logistic, Gompertz and Monomolecular models were considered in the study. All non-linear models were fitted to data by using Statistical software R. For validation of assumptions of residuals i.e., randomness and normality of residuals, Run's test and Shapiro wilk's tests were employed respectively while for goodness of fit and validation of models, Chi-square test and OSAF test were used. For getting best fitted models for forecasting yield, models are compared by ten different statistics R2, R2 7, R 28, RSS, MAPE, MAE, MSE, RMSE, RSE and MSE. nn. So, after analysing the data, Gompertz model is found better for Bilaspur district (Chhattisgarh) with FE% of 22.3 % which was found from actual and forecasted yield for year 2019-20, i.e., 0.76 (t/ha) and 0.93(t/ha) respectively. Forecasting of chickpea yield is made for next five years using fitted Gompertz model.

Yield Forecasting of pigeon pea in Karnataka using Auto-Regressive Integrated Moving Average Model (ARIMA)

Naveen, Mahesh Kumar and Nidhi
Dr. Rajendra Prasad Central Agricultural University, Pusa, Samastipur

The present study entitled examine “Yield Forecasting of pigeon pea in Karnataka using Auto-Regressive Integrated Moving Average Model (ARIMA)”is based on the ARIMA models for forecasting pigeonpea yield in Karnataka. The secondary data for the years 1979-1980 to 2020-2021 was retrieved from reliable websites like the Department of Economics and Statistics and India Agri Stat.The different ARIMA models are judged on the basis of Autocorrelation Function (ACF) and Partial autocorrelation Function (PACF) at various lags. The data from 1979-80 to 2018-19 are used for model building and from 2019-20 to 2020-21 used for successful cross-validation of the selected model on the basis of the absolute percentage error. The ARIMA models are fitted to the original time series data as well as the first difference data to check the stationarity. The possible ARIMA model are identified on the basis of significant coefficient of autoregressive and moving average components. The possible ARIMA model are identified on the basis of significant coefficient of autoregressive and moving average components. The best fitted models are selected on the basis of low value of Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), Mean Absolute Error(MAE) and BIC. Among all the models ARIMA (1,0,1) model is best fitted for pigeonpea yield in Karnataka. The best fitted ARIMA model has been used to forecast the yield for the year 2019-20 to 2023-24. The model showed the yield forecasting of pigeon pea for the year 2019-2020 to be about 683 kg/ha with upper and lower limit 333 and 1063 kg/ha respectively, with a forecasting error of 6.52%.

Study on multiple components for forecasting in price volatility for Agricultural commodities

Anushaka Garg, K.N. Singh, Achal Lama

ICAR – Indian Agricultural Statistics Research Institute, New Delhi

Agricultural commodity prices tend to be more volatile due to seasonality, inelastic demand and many more reasons. Increase in price volatility implies higher uncertainty about future prices, a fact that can affect producer’s welfare. Price volatility provides a measure of the possible variation or movement in the price variable.

The most widely used non-linear non-parametric generalized autoregressive conditional heteroscedasticity (GARCH) model did not allow for explanatory variables in the conditional variance. To overcome this limitation, recently the focus has shifted to multiple component models. In particular, the class of generalized autoregressive conditional heteroscedasticity mixed-data sampling (GARCH-MIDAS) models has been proven to be useful for analyzing the link between financial volatility and the macroeconomic environment. In GARCH-MIDAS model, a unit-variance GARCH component fluctuates around a time-varying long-term component which is a function of explanatory variables. By allowing for a mixed-frequency setting, this method fills the gap between daily volatility and low-frequency explanatory

variables. Thus, in this study an attempt has been made to develop appropriate GARCH-MIDAS model for forecasting volatility of the agricultural commodities by identifying the key explanatory variables. Where the real-life data has been taken from website: https://agmarknet.gov.in for the year 2015 to 2020.

Machine Learning and Deep Learning Techniques in E-Learning

Sowndarya, C.A.[1], Shashi Dahiya[2]

ICAR-IARI, New Delhi

ICAR-IASRI, New Delhi

E-learning has become the mandatory component of all educational institutions like schools, colleges, and universities in the world due to the pandemic crisis of COVID-19. It is one of the thrust areas identified by the Ministry of Electronics and Information Technology of India for imparting education using educational tools and communication media. It is the learning facilitated and supported by Information and Communication technologies (ICT). E-learning mode and the related tools provide a platform for enhanced learning, cost effective delivery, flexibility of learning at the convenience of the learner, uniform quality content delivery, re-usability of the content etc.Lots of student, faculty and course related data gets stored into the e-learning platform database. This data must be analyzed in order to extract useful information using Machine Learning and Deep Learning Techniques.These techniques are becoming increasingly important these days in order to find out useful patterns from the data and provide useful predictions beforehand.In e-learning, these techniques can be used to classify students based on their learning style or performance and to recommend courses to students using various classification algorithms.They can also be used to study and compare the student performance, as well as perform Sentiment Analysis to identify discouraged students using supervised machine learning classification algorithms.In this way, these techniques can be used in E-Learning Personalization and to predict student dropout in Digital Learning. In this paper, various E-Learning challenges and research opportunities using Machine Learning and Deep Learning are explored for future work in this area.

Price forecasting of Chilli (Green) - A Statistical Evaluation

Pal Vikash*[1], Darji V.B.[2]

[1] *Banaras Hindu University, Varanasi*

[2]*B.A. College of Agriculture, Anand Agricultural University, Anand*

Price forecasting of perishable products like vegetables have important implications to its stakeholders. A timely and reliable forecast of prices for different agricultural crops is highly required and need of a day. Forecasting of prices for agricultural commodities (like vegetables) remain difficult because they are influenced by many factors. The uncertainty of future price, production and consumption level makes agricultural market strategy and investment planning difficult. Perishability, price changes and seasonal nature of vegetables affect a lot to the Chilli (green) prices. In the present study, linear, quadratic and exponential trends were used for trend studies and forecasting of Chilli (green) prices. Also, ARIMA, ARCH/GARCH models and Artificial Neural Network (ANN) were employed for the study. For studying the error behaviour,

Jarque-Bera test was utilized. Statistical comparisons were made for different models using Root Mean Squared Error (RMSE) and Mean Absolute Per Cent Error (MAPE). Jarque-Bera test results showed that none of the model residuals followed normal distribution. In all, the comparison of different models tried in the study to forecast prices for Chilli (green) the Artificial Neural Network (ANN) model based on RMSE value performed better as compared to all the models studied.

Estimating Fisheries Reference Points using Time Series of Fish Catch and/or Fishing Effort: Application of Bayesian Approach

Sreepriya V.*, Judith Das,Somy Kuriakose, Eldho Varghese and J. Jayasankar
ICAR-Central Marine Fisheries Research Institute, Kochi

Estimation of fisheries reference points is essential to develop management strategies for sustainable harvest of marine fishery resources.A widely accepted approach towards this is to derive sustainable harvest levels using time series data on fish catch and fishing effort based on surplus production models that describe the biomass dynamics. In surplus production models, biomass is modelled as a function that incorporates species recruitment, growth, and natural mortality without taking into account the age or size structure of the population. Recently, the Bayesian methodology has been rigorously applied in developing stock assessment strategies.In this paper, a comparative analysis has been made among the three popular approaches viz., Catch-Maximum Sustainable Yield (CMSY), Bayesian state-space implementation of the Schaefer production Model (BSM) and Abundance Maximum Sustainable Yield (AMSY) for estimating the reference points for management of *Scomberomoruscommerson*(narrow-barred Spanish mackerel) using the time series of landings and fishing effort data collected from the South-West Coast of India.

Development of an AI and IoT Decision Support System for Sheep

Ambreen Hamadani and Nazir A. Ganai
Sher-e-Kashmir University of Agricultural Sciences and Technology of Kashmir

As the challenges of food insecurity and population explosion become more pressing, there is a dire need to revamp the existing breeding and animal management systems. This can

be done by the intensification of our farming systems and the improvement of the genetic merit of animals. A fundamental requirement for animal breeding is the availability of accurate and reliable pedigreed data and tools facilitating sophisticated computations. Keeping this in view, Smart Sheep Breeder (SSB) was developed. It is a fully functional multi-use online artificial intelligence and internet of things compatible decision support system (DSS). It is capable of automatic performance recording, farm data management, data mining, biometrical analysis, e-governance, and decision-making in sheep farms. A mega database thus developed would be capable of across-farm genetic ranking of sheep and effective dissemination of germplasm. The first-of-its-kind system in India is available as a web-based tool and android application which facilitates performance recording and generates customized reports on various aspects of sheep production. SSB uses artificial intelligence and biometrical genetic algorithms to calculate breeding values, and inbreeding coefficients, construct selection indices and generate pedigree, and history sheets as well as more than 40 types of custom-tailored animal and farm reports and graphs. The algorithms used were validated using farm data and comparison with established methods. Smart Sheep Breeder could thus prove to be indispensable for the present farming systems which could subsequently be used by breeders across India.

Evaluation of Machine Learning Algorithms for the Prediction of Genetic Merit

Ambreen Hamadani, Nazir A Ganai, S. Mudasir, S. Shanaz, S. Alam, I. Hussain
Sher-e-Kashmir University of Agricultural Sciences and Technology of Kashmir

As the amount of data on farms grows, it is important to evaluate the potential of artificial intelligence for making farming predictions. Considering all this, this study was undertaken to evaluate various machine learning (ML) algorithms using 52-year data for sheep. Data preparation was done before analysis. Breeding values were estimated using Best Linear Unbiased Prediction. 12 ML algorithms were evaluated for their ability to predict the breeding values. The variance inflation factor for all features selected through principal component analysis (PCA) was 1. The correlation coefficients between true and predicted values for artificial neural networks, Bayesian ridge regression, classification and regression trees, gradient boosting algorithm, K nearest neighbors, multivariate adaptive regression splines (MARS) algorithm, polynomial regression, principal component regression (PCR), random forests, support vector machines, XGBoost algorithm were 0.852, 0.742, 0.869, 0.915, 0.781, 0.746, 0.742, 0.746, 0.917, 0.777, 0.915 respectively for breeding value prediction. Random forests had the highest correlation coefficients. Among the prediction equations generated using OLS, the highest coefficient of determination was 0.569. A total of 12 machine learning models were developed from the prediction of breeding values in sheep in the present study. It may be said that machine learning techniques can perform predictions with reasonable

accuracies and can thus be viable alternatives to conventional strategies for breeding value prediction.

Genome-wide identification of copy number variation in equines

Nitesh Kumar Sharma[1], Anuradha Bhardwaj[2], Sarika Jaiswal[1], Mir Asif Iquebal[1], Dinesh Kumar[1], Anil Rai[1]
[1]ICAR-Indian Agricultural Statistics Research Institute, New Delhi
[2]ICAR-National Research Centre on Equines, Hisar

CNV (copy number variation) is a major source of genomic variation linked to phenotypic variation and disease vulnerability. In the present study, we used genomic data from 96 animals evaluated using the Axiom® Equine Genotyping Array (670,796 single-nucleotide polymorphisms (SNPs) with an average marker interval of 3.4 kb on the autosomes) to undertake a genome-wide autosomal screen to discover CNV in the six equine breeds, namely, Kathiawari, Marwari, Manipuri, Zanskari, Bhutia and Spiti. In 96 equine breed samples, 2668 CNVs and 381 CNV regions (CNVRs) were discovered after genotype quality check. The average size and median of CNVs were found to be 110.32 Kb and 49.91Kb, respectively. The coverage of CNVR was ~ 2.15% (49Mb/2280.94Mb) of equine autosomes genome sequence and 1.98% (49Mb/2474.93Mb) of equine whole genome. GO analysis was used to characterize the genes found in CNVRs. In GO analysis, the most over-represented molecular function was ola-factory receptor activity and the most overrepresented biological process is sensory perception. In equine breeds, the genome-wide CNVRs map will aid in the discovery of genetic variation and disease susceptibility alleles.

Genome wide Identification of Polymorphic SSR markers in Allium species

Princy Saini, Mir Asif Iquebal, Sarika Jaiswal, Anil Rai

ICAR-Indian Agricultural Statistics Research Institute, New Delhi

Allium species are very important due to their medicinal values. A. cepa and A. sativum are the most essential crop of the genus Allium. Microsatellites or simple sequence repeats (SSRs) are among the genetic markers commonly utilized in research. A total of 3429104 SSRs were identified in onion and 4053398 SSR in garlic. It is observed that the Dinucleotide SSRs (45.08%) in Onion and Dinucleotide SSRs (21.47%) in garlic were the most abundant, followed in descending order by mono-, tri-, tetra-, penta-, and hexanucleotide SSRs. The total length, density and frequency of SSRs were 76268248 bp, 5105.85 bp/Mb and 169.32 loci/Mb, respectively in Onion genome and for garlic genome the total length, density and frequency of SSRs were 9,05,19,672 bp, 6259.89 bp/Mb and 202.2107471 loci/Mb

respectively. We have categorized SSR loci into class I or hypervariable and class II or potentially variable markers on the basis of length of repeat motif which consist of SSRs >= 20 bp, and in between >= 12 bp and < 20bp SSRs, respectively. In dinucleotide, AT and TA were the most abundant SSR motifs in Onion and garlic genome respectively. In onion, the highest number of SSRs motif in chromosome 7 and in Garlic, chromosome 2 has the highest frequency of SSRs. This study will be useful for molecular breeding, varietal identification, and future strategy planning for germplasm conservation.