J. Ind. Soc. Agril. Statist.
62(1), 2008 : 96-107

Proceedings of the Symposium on "Statistical Aspects of Research in Agricultural and Environmental Sciences Status and Scope"

Chairman : Dr. V.K. Gupta
Convenors : Dr. Prajneshu
Prof. N.C. Das

Six papers covering various aspects related with the theme of the symposium were presented by the following speakers :

1. Statistical Aspects of Design and Analysis in Agricultural Engineering Research — Dilip Kumar Swain and Bimal Chandra Mal
2. Statistical Aspects of Research in Animal Sciences : Status and Scope — B. Singh
3. Small Area Estimation in Agriculture – Some Applications — U.C. Sud
4. Nonlinear Statistical Model for Characterizing-cum Growth of Bamboos — Gitasree Das, Arun Jyoti Nath, Subrata Nandy and Ashesh Kumar Das
5. Structural Time Series Modelling: Status and Scope — S. Ravichandran
6. Statistical Aspects related to the Registration of Plant Varieties in India — R.C. Agrawal
7. Model-based Inference in Finite Population Sampling Theory — B.V.S. Sisodia
8. Econometric Approach of Estimation of Technical Efficiency of Fish Culture Farms — A.K. Roy

The following recommendation emerged out based on the discussions during the symposium:

- The main emphasis was on discussing current status and scope of various areas of Agricultural Statistics, like Sample Surveys, Design of Experiments and Statistical Methodology to solving important research problems in various branches of Agricultural and Environmental Sciences.

ABSTRACTS OF THE PAPERS PRESENTED

1. Statistical Aspects of Design and Analysis in Agricultural Engineering Research

Dillip Kumar Swain and Bimal Chandra Mal

Statistics, as an essential tool for engineering and sciences research, has been greatly affected by the fields with which it interacts. Statistical intervention has improved the management of the natural resources such as agriculture, fishery, forestry, wild life, etc. The common features of statistical approach that takes agricultural engineering research beyond the traditional are: problem focused, management, system based and integrative. The research aims to solve the problems that are well defined with clear geographical boundary. Specific objectives are retained in the research to enable effective design of the study and analysis of the resulting data. The research usually aims to understand the impacts of changing resource management, and also to bring about improvements through statistical analysis. In Agricultural system, the key questions to be answered are generally expressed as a statement of hypothesis, which are usually suggested by past experience, observations, and at times by theoretical considerations. The hypothesis has to be verified or disproved through experimentation. The statistical design and analysis of some experiments of agricultural engineering research are discussed in this paper.

Selection of rice variety and evaluation of the crop management practices were the objectives to increase the rice production in non-water stressed rice-growing areas in Asia during the wet season (Swain *et al.* 2007).

Statistical analysis and crop simulation model were used for selection of rice variety and appropriate nutrient management strategy. A field experiment was conducted with 12 popular rice verities and varying N levels in a split-plot design, where varieties were in the main plot and N levels in sub-plots of the design. Grain yields and time series growth parameter data collected from the experiment were analyzed as per the design. Grain yield of these varieties was simulated at varying N levels using ORYZA 1N model. A paired 't' test confirmed that simulated grain yield of the varieties were in close approximation with the observed values under changing N levels. The rank order of each variety was determined from simulated and observed grain yield at optimum N application level (80 kg N/ha). Duncan's Multiple Range Test (DMRT) was used to find the values, which were significantly different. Through the DMRT, the rice variety Ranjit was selected. Optimum N level of 80 kg/ha was recommended through the statistical analysis of grain yield. The variety Ranjit with 80 kg N/ha application was recommended as the developed technology for rainfed non- water stressed regions. The performance of the technology (variety: Ranjit, and N level: 80 kg/ ha) was evaluated against farmers' practice (variety: Swarna, and N level: 40 kg/ha) in a Technology Verification Trial conducted in 2^3 factorial experiment design. The Technology verification trials of this practice produced 5.51 Mg ha^{-1} of rice, compared with 4.36 Mg ha^{-1} grown with the conventional practices of farmers.

Several researchers used statistical methods to evaluate the performance of crop simulation models. Jones and Kiniry (1986) used means and standard deviations for observed and simulated data, and linear regression parameters such as intercept (α), slope (β), and coefficient of determination (R^2). However, Thornton and Hansen (1996) concluded that use of the F-test (to test the null hypothesis that the regression line has unit slope and an intercept of zero) could be severely misleading. Willmott (1982) and Willmott *et al*. (1985) recommended the use of root mean square error (RMSE), and D-index (index of agreement), but suggested that RMSE is the best measure as it summarizes the mean difference in the units of observed and predicted values. The D-index is a descriptive measure, and can be applied to make cross-comparisons between models. Kobayashi and Salam (2000) used mean squared deviation (MSD), but concluded that root mean square deviation (RMSD) is probably more intuitive than MSD because the deviation is expressed on a relative basis against observations.

Timsina and Humphreys (2006) evaluated the performance of crop simulation model CERES-Rice and CERES-Wheat in rice-wheat system using statistical analysis. CERES-Rice predicted anthesis and maturity dates of rice fairly well with a low RMSE of 4-5% and high D-index of 0.94-0.95, but grain and biomass yield were more variable (RMSE = 23% for both; D-index 0.90 and 0.76, for grain and biomass, respectively). Similarly CERES-Wheat predicted the anthesis and maturity dates quite well (RMSE = 4–5%; D-index = 0.94–0.99), and grain and biomass yield reasonably well (RMSE =13-16%, and D-index = 0.86-0.97).

Designing of an efficient irrigation system requires information on infiltration characteristics of a soil, which is vital for both irrigated and rainfed agriculture. Infiltration was modeled and spatial variability of infiltration was analysed in a wasteland of Kharagpur, West Bengal, India (Machiwal *et al*. 2006). A total of 24 infiltration tests were conducted on a systematic squared grid pattern over the study area using double-ring infiltrometers. The observed infiltration data from all the test sites were fitted to four selected infiltration models and a best-fit model for individual sites was identified through statistical analysis. Based on the results of this study, it was concluded that the study area has wide spatial variability. Hydraulic parameters optimization techniques for production wells were compared using RMSE (Jha *et al.* 2004). Conventionally graphical step-drawdown test data analysis was compared with non-conventional optimization technique called Genetic Algorithm (GA) for the hydraulic parameters such as well loss coefficient and aquifer loss coefficient. It was revealed that the GA technique yielded more reliable well parameters with significantly low values of RMSE.

Agricultural systems and simulation analysis can be used to integrate the knowledge of bio-physical processes governing the soil-plant-atmosphere system to evaluate the production uncertainties associated with various management options. Statistical analysis coupled with simulation models predicts precise use of inputs i.e. water, nutrient, etc and evaluates management options for sustainability of food production in various agro-ecosystem which leads to precision agriculture.

Indian Institute of Technology, Kharagpur - 721 302

2. Statistical Aspects of Research in Animal Sciences: Status and Scope

B. Singh

The main statistical aspects of research in animal sciences are development and applications of experimental designs, methods of data analysis, linear models, sampling methodology, discriminant analysis, and fitting and predictive models for estimation and analysis of data in addition to creation of databases related to livestock statistics. Many advanced statistical techniques have been developed but they have not been used to the desired extent. There is enough scope for application of advanced statistical methodologies in various disciplines of animal sciences. Statistical methodologies and their aspects of research in animal sciences are reviewed and their status and scope are examined.

Designs and Analysis of Experiments

Three aspects of animal research, reduction of animal numbers, refinement of statistical methods and replacement of animals by non-animal alternatives (statistical designs and methods of analysis), should be considered while designing the animal experiment. CRD and RBD, the commonly used designs in animal experiments, are quite simple and robust. LSD is used to control the variation in two variables. The small order Latin squares have few degrees of freedom for error which can be increased by using repeated Latin squares. Factorial designs are used to evaluate two or more factors, simultaneously. The main advantage of factorial designs is the detection of interactions. The incomplete block designs are not frequently used in animal experiments.

To reduce the number of animals and the error variation the same experimental unit may be used for different treatments in different periods. Such designs are known by different names such as cross-over, change-over or switch-over designs. The disadvantage of such designs is the residual effects. One way to get rid of residual effects is the use of sufficient washout period which is not always possible due to limited time period. So the provisions should be made to estimate direct and residual effects, separately. The changeover designs balanced for first residual effects and with number of periods less than number of treatments has been introduced. The direct and residual effects of treatments can not be independently estimated in such designs except when an extra period is added to the basic design.

In many animal experiments the response variable is measured on the same animal under same treatment across occasions. Such observations will usually be correlated. A commonly used method for analysis of repeated measures designs is to compare treatment groups at each time point, separately. The other approach is to use summary measures. These approaches do not provide tests for time-point and interaction effects. The repeated measures ANOVA (RANOVA) provides tests for time points and interaction in addition to treatment effects. This method is valid under compound symmetry (CS) of covariance matrix. An approximate test may be carried out by adjusting degrees of freedom of mean squares by epsilon based on covariance matrix when CS condition does not hold. An alternative is to use multivariate approach that can not be applied in case of missing data. RANOVA for incomplete measurements has been provided by Singh (1997). The mixed model methodology analyzes such data correctly and efficiently. The results of RANOVA and mixed model methodology are very similar under CS and they differ only because of the incomplete data.

Most commonly used method for analysis of categorical data is chi-square method. The other method is to assign consecutive integer values to categories and then use ANOVA method. Recent methods include weighted least squares (WLS), generalized linear models and CATANOVA. The WLS methodology may be inefficient for small frequencies and can not be used with continuous covariates.

LSD and DMRT tests are too liberal and Tukey's and Bonferroni tests are conservative for multiple comparisons. Dunnett test is used for comparison of each treatment with control.

Applications of Linear Models

In fixed effects models, the method of least squares provides unbiased estimates of parameters. In mixed models the generalized least squares (GLS) estimator is best linear unbiased estimator for fixed effects when covariance matrix is known and it is unbiased in case of unknown covariance matrix for a wide class of variance components estimators.

Hofer (1998) gives a brief outline of the prediction of breeding values when variance components are known and discusses the consequences of using their estimates. The empirical BLUP, using estimates of variance components from a wide class of variance components estimators leads to unbiased predictors. The fixed and random effects parameters can also be estimated simultaneously by using the Henderson's mixed model equations.

For balanced data, ANOVA estimators of variance components are minimum variance quadratic unbiased and minimum variance unbiased under normality. For unbalanced data Henderson's methods are commonly used. The ML estimation has attractive large sample properties such as asymptotically unbiased, asymptotically normally distributed and asymptotically efficient. The complete least-squares analysis provides the most efficient estimates of variance components. The testing of variance components is simple in balanced random models and no exact testing for main class variance component in unbalanced two-way random model is available except under last stage uniformity. Approximate tests for main class variance component under general two-way unbalanced random model have been developed by several workers.

Discriminant Analysis

The Fisher's linear discriminant function (LDF) is commonly used in animal research. It yields optimal results if parameters are known and assumptions of multivariate normality and identical covariance matrices hold. Probability of misclassification (PMC) for Fisher's LDF is available in books and it is given by Singh (2001a) when parameters are unknown. The quadratic discriminant function (QDF) is optimal procedure in case of unequal dispersion matrices. PMC of QDF is given by Gilbert (1969) for special and by Singh (2001b) for general situations. LDF performs better in populations without reversals and otherwise the full multinomial procedure for binary data. In case of mixed-continuous and discrete variables location model approach has been suggested by Krzanowski (1980). A coupling procedure by combining an adequate number of discriminators for different types of data has been proposed by Wernecke *et al.* (1986).

Indian Veterinary Research Institute, Izatnagar - 243 122

3. Small Area Estimation in Agriculture –Some Applications

U.C. Sud

Statistics relating to various parameters of interest such as average yield, total size, important demographic parameters like birth, death etc. economic activities, education, health and natural resources are often collected by local administrative organization. Such local area statistics cannot be entirely relied upon when the decision makers need reliable data for formulating policies for distribution of funds among the lower administrative units for planning infrastructure development for creating basic amenities and for initiation of welfare projects for the local areas. In this context there arises a need for a mechanism for collecting reliable statistics at the local level.

Censuses, administrative files and sample surveys are three different mechanisms for generating data on various parameters. The data collected out of census and administrative files suffer from certain drawbacks like quality, lack of completeness etc. Seen in this context, sample surveys, if properly planned, are an important mechanism for data generation on various items of interest. Sample surveys are planned in such a way that they provide reliable estimates for larger areas. Thus, in the context of estimation of yield rates of crops the surveys are planned to provide reliable estimates at the district level. If from a sample survey planned for a larger area one is aiming to generate estimates for the lower level say a block or a gram panchayat there may be a situation where the data points falling in the small area may be too little. This may result in estimates with very large standard errors as to be of any use. In the worst case a particular small area may not have any data points. In this case it will not be possible to provide any estimates for the particular small area.

Faced with this scenario one is left with two choices. Either plan sample surveys for each of the local/small areas or use some small area estimation technique for estimation of parameters at the local area level. For instance, in the context of agricultural surveys, the random sampling crop cutting surveys proposed by the ICAR are being used to provide reliable estimates of crop yield at the district level. These estimates are obtained out of the General Crop Estimation Surveys (GCES) scheme organized under the administrative control of Directorate of Economics and Statistics (DES). With the emphasis on micro level planning estimates such as these are being demanded at the lower level which may be a tehsil or even a gram panchayat. The National Agricultural Insurance Scheme launched by the Government of India for the benefit of the farmers is a case in point. The present crop cutting experiments approach for framing estimates at the village panchayat level or even tehsil level may not be practicable in this situation as this will require conducting many more crop cutting experiments over and above the experiments that are already being conducted. The financial implications and organizations of so many crop cutting experiments will be very demanding. Therefore a small area estimation technique is needed in this case. The basic rationale of the approach of small area estimation is that certain assumption/models are conceptualized which are assumed to hold good for small as well as large area.

Early developments in Small Area Estimation Techniques were in the field of demographic studies. These were categorized under the general heading of Symptomatic Accounting Techniques (SAT). These techniques essentially involve utilizing current data from administrative registers and the related data from latest census. The synthetic method of estimation was developed by National Centre for Health Statistics (NCHS) in the year 1968. The method essentially involves using estimates for the larger area to borrow strength from outside the domain of interest to frame estimates for the local areas. Composite estimators which are the weighted combination of the synthetic estimator and the direct estimator are widely used for generating estimates for the smaller areas.

Using the small area estimation techniques, reliable estimates of milk production at the district level have been developed at the IASRI. For this purpose the synthetic estimator was developed using the breed of the animal as the grouping variable for developing the estimates at the district level. Another small area related carried out at the IASRI involved using satellite data for the Rohtak district along with GCES data on crop yield to develop synthetic estimator which in turn is used to estimate crop production in the different tehsils of Rohtak district. In this case the synthetic estimator was developed by using Normalised Difference Vegetative Index (NDVI) as the grouping variable for framing precise tehsil level estimates.

Using the farmers' inquiry based data on crop production and the GCES data on crop yield an estimator has been developed for framing estimates at the Gram Panchayat level. The approach is akin to small area approach in the sense that Block/District level estimates are scaled down to GP level estimates under the assumption that over estimation and under estimation are behaving in similar way in the entire area. Using the data on wheat crop for Basti district for the year 2000, estimates of wheat yield at the Gram Panchayat level were developed by Sharma *et al.* (2004) along with Percentage Root Mean Square Error (RMSE) by following the Small Area Crop Estimation Methodology (SACEM) approach. The results of the study indicate that the RMSE varied between 1-10%.

The above mentioned approach assumes that the gram panchayat level crop area figures are available. This may be true for temporary settled states but such figures are not available for permanently settled states like Orissa, West Bengal and Kerala. For these states area figures are generated using sample survey approach. Thus, every year 20% of the total cropped area is enumerated. The entire crop area is covered in five years. Needless to say that unlike for temporary settled states where Gram Panchayat wise area figures are available, for such states the figures at best are available for only 20% area. In this case estimates for the non-sampled area can be framed using the area level model. The area level models are basically mixed models wherein the direct estimators for the small areas are related to area-specific auxiliary data (assumed fixed) and area specific random effects through a linear model. Such models have been extensively used in the context of small area estimation studies. Use of these models and other related aspects along with results of analysis of data will be dealt with in detail in the lecture.

Indian Agricultural Statistics Research Insitute,
New Delhi - 110 012

4. Nonlinear Statistical Model for Characterizing Culm Growth of Bamboos

Gitasree Das[1], Arun Jyoti Nath[2], Subrata Nandy[2] and Ashesh Kumar Das[2]

The study on growth pattern of bamboos, especially the commercially most important parameter viz. the culm height, is important in developing scientific management systems for optimum yield. In this article nonlinear statistical models were used to characterize the culm height growth of two important bamboo species. One of these is an important non-timber forest product (NTFP) of North-East India known as *Melocanna baccifera* or Muli bamboo. The other one, *Bambusa cacharensis*, locally known as betua, is an important village bamboo species in the Barak Valley region of Assam. Villagers prefer to grow this species in their home gardens because of its multiple uses and having straight culm, preference being as high as 70 % compared to other species growing in this region.

The adequacy of the models was judged by testing the validity of the assumptions of randomness and normality of residuals by using one-sample run test and Shapiro-Wilk test respectively. Durbin-Watson test was used to examine the presence or absence of autocorrelation. To assess the goodness of fit of the suggested models, statistics viz. R^2, root mean square error (RMSE), mean absolute error (MAE) etc. were computed for each model. Gompertz model with additive and AR (1) error structure is found to be the most suitable model for describing the culm growth of *M. baccifera*. The physical interpretation of the parameters is also being discussed. The estimated parameters can be treated as a summary of the growth pattern and can be used to characterize the growth of the bamboo species. These values may be used for comparison across species or varieties or same species growing under different environmental conditions.

Gompertz model with additive and AR(1) error structure appears to be the most suitable model for the height growth data of B. cacharensis corresponding to clumps of different age groups viz. 2, 5, 10, 15 and 40 years. Wilk's-Test was used to test the equality of mean vectors (k, q, r, a) corresponding to clumps of different age groups. Next, Hotelling T^2 statistic was computed to test equality of two mean vectors (k, q, r, a) corresponding to clumps of different age groups**.** The constant k is known as the 'carrying capacity' of the environment. The growth curve increases to an upper limit k when t is large. Thus k is the most important parameter showing the maximum height that a bamboo culm can attain. Therefore one-way analysis of variance of k-values for clumps of different age groups was performed.

The Gompertz model with additive and AR(1) structure is identified by four parameters: k, q, r and a. Das *et al*. (2006) have shown that for M. baccifera $k = 1698.855$, $q = 10.839$, $r = 0.182$ and $a = 0.833$. The parameter q gives the position of the point of inflection i.e. the time when the growth rate is maximum. For M. baccifera estimated maximum growth rate $rk / e = 113.75$ cm per week. The growth attained at this point is estimated as $f = k / e = 624.97$ cm. And a is the autoregressive model parameter. Hence the observed culm height growth curve of M. baccifera may be characterized by its parameter estimates as (1698.855, 10.839, 0.182, 0.833). The role of k, q, r and a is very important in developing scientific management systems for optimum yield of bamboo.

For B. cacharensis the estimated parameters can be used to characterize the growth pattern of bamboo culms for clumps of different age groups. Bamboo clumps of different age groups can then be compared on the basis of these parameters. Each growth curve may be summarized by its parameter estimates as a single low-dimensional multivariate observation. Gompertz model with additive and AR(1) error structure was fitted to individual culms from clumps of different age groups. Wilk's-Test was used to test the equality of mean vectors (k, q, r, a) corresponding to clumps of different age groups. Since $P(F(8, 82) > 6.9956) = 5.351E\text{-}07$, clumps of different age groups are significantly different with respect to their growth parameters.

Hotelling T^2 statistic was computed to test equality of two mean vectors (k, q, r, a) corresponding to clumps of different age groups. A perusals the p-values confirms that 10-year old clumps are significantly different from the rest in terms of the parameters k, q, r, a. Further, one-way analysis of variance of k-values for clumps of different age groups indicates ($p < 0.001$) best culm growth for 10-year old clumps. However, it is worth mentioning here that further investigation is needed to confirm this fact by considering clumps of some more age groups between 15 and 40 years. Thus while developing management strategies, both age of culm as well as clump needs to be considered. However detail discussion on the same is beyond the scope of this work.

Thus the parameters (k, q, r, a) can be used to characterize the culm height growth of B. cacharensis growing in clumps of different age group. A perusal of the (k, q, r, a) values for M. baccifera (Das *et al.* 2006) reveals that these are distinctly different from the growth parameters corresponding to clumps of different age groups for B. cacharensis. Thus, the parameters can also be used to characterize the growth of different species. The parameter q gives the position of the point of inflection i.e. the time when the growth rate is maximum. For B. cacharensis this period varies between 6 to 8 weeks, whereas for M. baccifera the same is attained at almost 11^{th} week. For M. baccifera estimated maximum growth rate rk / e = 113.75 cm per week, whereas for B. cacharensis it varies from 132.82 cm to 141.83 cm per week. Finally the growth attained by M. baccifera at this point is estimated as f = k / e = 624.974 cm and the same for B. cacharensis varies between 334.16 cm and 411.11 cm. Lastly, maximum height that a culm of M. baccifera can attain is nearly 17 meters whereas that of B. cacharensis varies between 9 to 11 meters. Such comparisons are important in developing proper scientific management systems for optimum yield.

[1] *North-Eastern Hill University, Shillong – 793022*
[2] *Assam University, Silchar – 788011*

5. Structural Time Series Modelling: Status and Scope

S. Ravichandran

Box-Jenkins Auto Regressive Integrated Moving Average (ARIMA) time series technique is widely used for modelling time series data. However, this methodology can be applied only when the given data is stationary. Another limitation of ARIMA approach is that it is empirical in nature and so it does not provide any insight into the underlying mechanism. An alternative mechanistic and quite promising approach viz. "Structural time series modelling (STM)" approach can be effectively utilised for modelling time series data in the presence of various characteristics such as level, trend, seasonal, and cyclical fluctuations. In this approach, characteristics of the data dictate particular type of model to be adopted from the family and is much more wider than ARIMA technique.

STM models are formulated in such a way that their components are stochastic and are driven by random disturbances. The key to analysing of these models is the "State space" form, with state of the system representing Various unobserved components, such as trends, cycles and seasonals. Once in state space form (SSF), Kalman Filter provides means of updating the state, as new observations become available. Predictions are made by extrapolating these components into the future. Kalman filter, a smoothing algorithm is used for obtaining best estimate of the state at any point within the sample. Prediction and smoothing is carried out once parameters governing stochastic movements of state variables have been estimated. Estimation of these parameters, known as "hyperparameters", is based on Kalman filter. Once a model is estimated, its suitability can be assessed using goodness fit statistics, like Akaike information Criterion (AIC) and Schwartz- Bayesian Information Criterion (SBC). Lower the values of these statistics, better is the fitted model. STM models can be fitted to the data using "Structural Time Series Analyser, Modeller and Predictor (STAMP) software package, SsfPack2.2 software package or by SAS (Statistical Analysis System) Proc UCM.

In the absence of seasonal and cyclical components, the STM reduces to a Local level model (LLM). The estimation of hyperparameters and likelihood function can be evaluated by Kalman filtering via prediction error decomposition. Once hyperparameters are estimated, one-step-ahead prediction of the estimator and Prediction error variance are evaluated recursively by Kalman filter. Reduced form of LLM is ARIMA (0, 1,1) model. Local linear trend model (LLTM) is also obtained from STM, when there is presence of trend and error and when both the level and slope are stochastic. Hyperparameters are estimated using maximum likelihood method for state space models. Reduced form of a LLTM is ARIMA (0, 2, 2) model. Forecast function of LLTM performs better than that of corresponding ARIMA model. Local linear trend model with intervention effect (LLTMI) incorporates the impact of intervention, which is concerned with making inference about effects of known events. These effects are measured by including intervention, or dummy variables. As for LLTM, estimation of state vector for LLTMI is carried out by putting the model in state space form and applying Kalman filter recursively by treating weight as an explanatory variable. These models were applied for modelling and forecasting of country's sunflower yield time series data. It was satisfying to conclude that Technological Mission on Oilseeds set up by Government of India in 1986 had a significant positive impact in enhancing sunflower yield.

When Cyclical Fluctuations are prominent, the three important sub-models of STM are:

(i) ***Cycle Plus Noise Model (CNM),*** wherein the trend is assumed to be constant. For estimation of parameters, model has to be put in state space form and then Kalman filter, prediction and smoothing are applied. After estimation of parameters, prediction and smoothing are performed. The reduced form from ARIMA family corresponding to CNM is "Constant + ARIMA (2,2)".

(ii) ***Trend Plus Cycle Model (TCM),*** which has three sets of equations. Estimation of state vector and hyperparameters is carried out by putting the model in state space form and subsequently Kalman filter is applied with proper initial values. Once the parameters are estimated using prediction error decomposition, Kalman filter, prediction and smoothing can be applied.

(iii) ***Cyclical Trend Model (CTM),*** in which the cycle is actually incorporated within trend. The essential difference between TCM and CTM is that, in the former, the observation depends on cyclical fluctuations explicitly whereas, in the latter, it does so on the previous observation implicitly through the trend. ARIMA (2, 2, 4) model is the corresponding analogue of both TCM and CTM.

Application of above models to all-India annual lac production data having prominent cyclical fluctuations would be thoroughly described in the talk.

For describing Seasonal Fluctuations, two important STM models are:

(i) ***BSM with dummy seasonal (BSMDS):*** Here, seasonality parameter is used in dummy variable form. For estimation of parameters equations have to be first put in state space form. Once BSMDS is in state space form, Kalman filter is performed to obtain updating, prediction and smoothing of state vector. Prediction and smoothing are carried out only when the hyperparameters are estimated using EM algorithm.

(ii) In ***BSM with trigonometric seasonality (BSMTS),*** the seasonality parameter is described in terms of "trigonometric seasonal". Further, if the trigonometric seasonal components are incorporated into a linear trend plus error model and the full set of trigonometric terms is included, the two models, viz. BSMDS and BSMTS will provide identical results.

Details of these models for modelling and forecasting of quarterly landings in Tamil Nadu in respect of two fish species, viz. silverbellies and croakers would be given in the talk.

Scope of Future Research Work

(i) In the above, we have considered only "Linear state space models". Nonlinear state space models may be investigated by employing advance techniques, such as Extended Kalman filter and Generalized Kalman filter. Subsequently, software for fitting such models also needs to be developed.

(ii) The above Univariate models may be extended to deal with bivariate situations, leading to "Bivariate structural time series models". These can then be applied for modelling and forecasting of say, rice production and rainfall time series data simultaneously.

(iii) In reality, there are multi – species present in any subsystem. For example, in fisheries and entomology, there are interacting species having prey-predator, competition, or symbiotic types of interactions. Appropriate estimation procedures along with relevant computer programs to handle such situations have to be developed in order to apply "Multivariate structural time series models" to real-life data.

Directorate of Rice Research, Hyderabad - 500 030

6. Statistical Aspects Related to the Registration of Plant Varieties in India

R.C. Agrawal

As per "Protection of Plant Varieties and Farmers' Rights (PPV&FR) Act 2001", a new variety must conform to the criteria of Distinctiveness, Uniformity and Stability (DUS) in suitable tests before it is eligible for protection. According to Section 15(1) of the PPV & FR Act 2001, "A new variety shall be registered under this Act if it conforms to the criteria of novelty, distinctiveness, uniformity and stabilitys. Section 15(b) defines a variety to be distinct, if it is clearly distinguishable by at least one essential characteristic from any other variety whose existence is a matter of common knowledge in any country at the time of filing of the application. Section 15(c) defines a variety to be uniform, if subject to the variation that may be expected from the particular features of its propagation it is sufficiently uniform in its essential characteristics. A variety is called stable, if its essential characteristics remain unchanged after repeated propagation or, in the case of a particular cycle of propagation, at the end of each such cycle.

According to the Section 29 PPV & FR Rules, 2003, "the special tests shall be conducted only when DUS testing fails to establish the requirement of distinctiveness" and "The DUS testing shall be field and multi-location based for at least two crop seasons and special tests be laboratory based". Further, the "DUS test shall be necessary for all new varieties except essentially derived variety" and "the DUS test shall be conducted on a minimum of two locations".

The design of the growing trial or other tests, with regard to aspects such as the number of growing cycles, layout of the trial, number of plants to be examined and method of observation, is largely determined by the nature of the variety to be examined. Guidance on design is a key function of the Test Guidelines.

Three Types of Characteristics in DUS Trails

1. "Qualitative characteristics" are those that are expressed in discontinuous states (e.g. sex of plant: dioecious female (1), dioecious male (2), monoecious unisexual (3), monoecious hermaphrodite. These states are self-explanatory and independently meaningful. All states are necessary to describe the full range of the characteristic, and every form of expression can be described by a single state. The order of states is not important. As a rule, the characteristics are not influenced by environment.

2. "Quantitative characteristics" are those where the expression covers the full range of variation from one extreme to the other. The expression can be recorded on a one-dimensional, continuous or discrete, linear scale. The range of expression is divided into a number of states for the purpose of description (e.g. length of stem: very short (1), short (3), medium (5), long (7), very long (9)). The division seeks to provide, as far as is practical, an even distribution across the scale.

   The Test Guidelines do not specify the difference needed for distinctiveness. The states of expression should, however, be meaningful for DUS assessment.

3. In "Pseudo-qualitative characteristics" the range of expression is at least partly continuous, but varies in more than one dimension (e.g. shape: ovate (1), elliptic (2), circular (3), obovate (4)) and cannot be adequately described by just defining two ends of a linear range. In a similar way to qualitative (discontinuous) characteristics – hence the term "pseudo-qualitative" – each individual state of expression needs to be identified to adequately describe the range of the characteristic.

Determining whether a difference between two varieties is clear depends on many factors, and should consider, in particular, the type of expression of the characteristic being examined, i.e. whether it is expressed in a qualitative, quantitative, or pseudo-qualitative manner.

National Bureau of Plant Genetic Resources, Pusa Campus, New Delhi - 110 012

7. Model-Based Inference in Finite Population Sampling Theory

B.V.S. Sisodia

Conventional approach to sample survey theory is based on probability sampling for both sample selection and inference from the sample data, which is referred to as design-based approach. Plausible population models have been used at the design stage of a survey to help choose good sampling designs and estimators, but the inferences remained model-free (Cochran 1963, pp. 214, among others). In early seventies, Royall (1970-71) advanced a model-dependent approach in which population is assumed to obey a specified model, and model distribution provides valid inference about the finite population parameters like total or mean for a given particular sample, no matter in whatever manner it is selected.

A hybrid approach called model-assisted approach which uses supplementary information through population models but inferences are design-based. Sarndal *et al*. (1992) have given a detail account of this approach. It has been found that Generalized Regression Estimators play a prominent role in this approach.

The inferences in model-based and model-assisted approaches could be sensitive to model misspecification. Therefore, a lot of works have been advanced towards the robustness of the estimators under these approaches (Royall 1973a, 1973b; Scott *et al.* 1978; Write 1983; Tam 1987; Singh *et al.* 1987; Royall 1992; Dorfman and Valliant 2000; Kaushal and Sisodia 2007, among others). It has been shown that most of the estimators are robust under a concept of balanced sampling, overbalanced sampling and weighted balanced sampling. These balanced sampling procedures are based on moments of the auxiliary variables used in the models.

The works on robustness in stratified sampling are further advanced by Kaushal and Sisodia (2007) when the regression coefficients in population model are common across the strata. Very useful results are obtained by them.

Shrinkage estimators have also been developed under super population models following the concept of Thompson (1968). This work has been further extended to stratified sampling and the robustness of the shrinkage estimators has been worked out under the possibilities of varying and common regression coefficients in the model across the strata. Most prominent contributors in this area are Bolfarine (1986), Bouza (1990, 1994), Bouza and Alende (1991), Sisodia and Kaushal (2007) and Kaushal and Sisodia (2007).

An attempt has been made in the present article to review the most important works on model-dependent and model-assisted approach of inference in finite population sampling theory. Future scope in this area of research work is also highlighted.

Narendra Deva University of Agriculture & Technology, Faizabad-224 229

8. Econometric Approach of Estimation of Technical Efficiency of Fish Culture Farms

A.K. Roy

Aquaculture plays a vital role in ensuring fish availability for food security and nutrition in India. This sector has to develop and grow in an economically viable and environmentally sustainable fashion. That is possible through improved water management, better feeding strategies, genetic improvement of cultivable species and better health management leading to enhanced productive efficiency of aquaculture at farm level. Therefore measuring Technical Efficiency (TE) at farm level and identifying factors associated with efficient production and assessing potential for and sources of future improvements are need-of-the-hour for establishing sustainable aquaculture. Instead of increasing the use of inputs to increase production, efforts should be made towards output growth through improved TE, which means to produce more by utilizing inputs at hand more efficiently. Keeping these points in view, an attempt was made to estimate the technical efficiency of 221 aquaculture farms of Kolleru lake area, which is known as the Carp Pocket of India, located at Krishna-Godavari Delta in South India producing an estimated 70,000 tones of fish per year.

A stochastic frontier production function model was utilized to estimate the technical efficiency of carp farms using a software FRONTIER with the objective of identifying the role of various input on fish production as well as to assess the impact of farm specific and socio economic variables on technical efficiency of aquaculture farms. The estimated mean technical efficiency (TE) was observed to be 0.7260. The highest significant elasticity of coefficient was demonstrated for feed (0.2001) followed by organic manure (0.1411) justifying the importance of these two inputs in yield of carps. Assuming that socio-economic, demographic, farm specific, environmental and non-physical factors are likely to affect the efficiency of operation of any farm, in the present study altogether twenty-five nonphysical socio-economic, demographic and farm specific variables were included in the model to test the possible impact and association with the technical efficiency of aquaculture farms operating in the Kolleru lake area. Of the twenty-five variables considered for technical efficiency model, socio economic factors like religion, age, education, primary occupation, pond size, ownership type, source of water, depth of water, sources of advice taken, method of application of drugs, calamities faced, types of labour, types of feed and disease encountered have a positive impact on technical efficiency. On the other hand caste, number of children, experience, renovation, asset cost, source of seed, loan, harvesting technique, method of application of feed, periodical netting for biomass checking and types of organic manure showed a negative association with technical efficiency. Pond size and source of water were found significant at ($p < 0.05$) and types of labour, religion, method of application of drug, primary occupation and calamities are found significant at ($p < 0.10$). Likewise of the variables showing negative impact, periodical netting for biomass checking is found highly significant ($p < 0.05$), source of seed and harvesting technique and caste were found significant at lower level of significance ($p < 0.10$) demonstrating the influence of these factors in efficiency of carp culture operation**.** In-depth analysis of impact of pond size on TE indicates that a consistent increase in mean TE with the size of the farm showing highest TE value of 0.7552 for aquaculture farms of size greater that 10 ha.This may be due to the fact that larger farms are capable to capture the economics of size and operates at higher efficiency levels compared to those of marginal and small farms. Pond area is reported to have positive impact indicating large operations are technically more efficient than smaller ones. Pond size is reported to have positive relationship on productivity. Water is the primary component of all aquacultural activities and its importance has been reflected through significant positive impact on TE. Farmers those who have set up their farms by the side of irrigation canals facilitating regular intake and periodical discharge of water were found to operate with significantly higher mean technical efficiency of 0.7287 compared with those using water from other sources with a mean TE of 0.6648 ($F_{cal} = 5.57 > F_{crit} = 3.88$).

Types of labour employed also showed positive impact on TE. Farms employing permanent labour could utilize their services for timely operation of various farm activities reflecting significantly higher mean TE of 0.7354 compared with those who employed temporary labour which a mean TE of 0.6957 ($F_{cal} = 10.11 > F_{crit} = 3.88$). The majority of farms were operated by Hindu's who are influential and economically sound having an advantageous position for securing resources for investment resulting in significantly higher TE of 0.7275 compared to 0.6207 shown by other religions ($F_{ca\ l} = 5.33 > F_{crit} = 3.88$). The method of application of drug has also shown a positive impact over TE. To combat disease, drugs were used either diluting or spraying in water or mixed with feed shown an estimated TE of 0.7292, 0.7237 and 0.6284 respectively ($F_{cal} = 1.62 < F_{crit} = 3.04$). Primary occupation was also found to have a significant role indicating that farmers doing fish farming as the main activity showed a significantly higher mean of 0.7359 compared to those of others engaged in other activities, showing a mean TE of 0.6896 ($F_{cal} = 13.02 > F_{crit} = 3.88$). Impact of calamity on TE was found significant ($p < 0.10$). One possible explanation may be that Kolleru Lake is located in the flood and cyclone prone areas of Krishna and Godavari districts of Andhra Pradesh. As a result farms that were affected by flood water are automatically flushed out of organic load and metabolic wastes accumulated due to regular use of organic inputs. Farms reported to have been affected by calamity showed higher mean TE of 0.7276 compared to a mean TE of 0.7210 demonstrated by farms not affected by calamities during the last five years of operation, although the difference is not statistically significant ($F_{cal} = 5.57 < F_{crit} = 3.88$).

To sum up, the empirical results suggest that there are significant possibilities to increase efficiency levels by increasing pond size, provision for supply of fresh water, right method of application of drugs, employing labour on regular basis and leasing out the water bodies to fish farmers and educating the farmers to procure seed from reliable sources, proper harvesting technique and avoiding periodical netting. Analysis of potential TE improvement of aquaculture farms reveals that if the average farmer in the sample achieves the highest TE level that farmer would save 26.7 per cent in cost. The most technically inefficient farmer on the other hand would save 48.9 per cent. The study suggests that with the same level of input, there is potential to enhance the yield by 27.40% through efficient use of resources and improvement of technical efficiency at farm level from Kolleru Lake area. Differences in efficiency levels that are identified and explained with respect to various farm specific variables having impact on TE through estimating a model for technical inefficiency effects will help future researchers, planners and farmers in efficient management of their farms resulting in optimum utilization of resources.

Central Institute of Freshwater Aquaculture, Kausalyaganga, Bhubaneswar-751002