



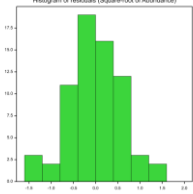






**DRAFT**  
**Biometrics and Statistics Section**  
**DDG-R**

**ICARDA's Strategy for Biometrics and Statistics Support to its Research:**  
Enhancing Research Quality for Innovation and Knowledge Sharing in Agri-Food  
Systems

<p>Barley, Rabat Station-March 2015</p> 	<p>Chickpea</p> 	<p>Lentil</p> 
<p>Faba bean</p> 	<p><math>\underline{Y} \sim N(\underline{X}\underline{\beta}, \sigma^2)</math></p> <p>sqryTB</p> <p>Histogram of residuals (Square-root of Abundance)</p> 	<p>Wheat, Rabat Station-March 2015</p> 
<p>Olive groves on steep slopes, Bitiya, Syria</p>  <p>Figure 3: An overview of Bitiya area; olive groves on steep slopes</p>	<p>Alley cropping</p> 	<p>Training</p> 

**Keywords:** Biometrics, Bioinformatics, Statistics, Statistical Computing and Experimental Research Databases

## **ICARDA's Strategy for Biometrics and Statistics Support to its Research:**

Enhancing Research Quality for Innovation and Knowledge Sharing in Agri-Food Systems

### Contents

0. Executive summary
1. Goals and objectives
2. Current status overview
3. New Challenges
4. Organization
5. Workplan 2016-2020
6. Annexures

### **1. Executive summary:**

This document describes the goals and objectives of Biometrics and Statistics Section (BSS), highlights current status and the challenges in the changing agricultural landscape to sustainably support the livelihood and the environment. BSS aims to provide biometric and statistics support to programs on planning of experiments and surveys, analysis of data, drawing statistical inferences, presentation of results, and statistical review of scientific documents. It also shares the relevant techniques and tools with the NARS scientists through training courses. BSS devises new methodologies for specialized situations if not available. Another area of expertise is to develop statistical computing modules and share them with ICARDA and NARS scientists. This effort leads to enhance the reliability, efficiency, repeatability and interpretability of experimental results leading to high level of science quality. It lists a number of key statistical issues and ongoing activities. The clusters of activities mentioned in the preproposals for 2<sup>nd</sup> Call of CRP DCLAS and Wheat have also been listed for easy harmonization of ICARDA BSS activities.

### **1. Goals and objectives**

#### Goals:

To enhance the reliability, efficiency, repeatability and interpretability of experimental results leading to high level of science quality.

#### **Objectives:**

1. The main objective of the Biometrics and Statistics Section (BSS) is to provide biometric and statistics support to programs on planning of experiments and surveys, analysis of data, drawing statistical inferences, presentation of results, and statistical

review of scientific documents and to devise specialized applications of advanced techniques required by the research programs.

2. To automate the scientific computing to minimize the cost
3. The upgrading the skills of the NARS scientists in Biometrics and Statistics
4. To store, manipulate, retrieve, analyse and interpret experimental data a) on germplasm including those in ICIS (molecular marker & phenotype data) to exploit for selection of desired genetic material, b) in International Nursery Data Management and Analysis and SEEDMAN for seed management, c) cropping systems for their productivity and sustainability studies, and) from Water, Land and Ecosystem research

## **2. Current Status of the Support**

Biometrics and Statistics Section (BSS) currently supports and contributes to enhancing the quality of research by advising efficient protocols for experimentations and surveys, sound statistical methods for data analysis and drawing statistical inferences, presentation of results, and conducting statistical review of scientific documents. Experimental designs often used for field trials include complete blocks and alpha-designs for variety trials, factorial experiments, phased entry in agronomic and long-term rotation trials, and specific design constructed for multi-disciplinary and contexts research. Statistical analyses of data are carried out using various techniques including analysis of variance and means, residual plot analysis, general and generalized linear models fitting, errors-in-variables, multi-variate methods, methods for spatial variability and estimation of variance components. More recently we are adapting Bayesian approach to integrate current data with prior information available during the course of experiment.

A number of new biometric techniques to meet specific needs of the research projects were developed for both germplasm enhancement and natural resources management. These include, for example, analyses of yield trials with common checks, an index of measuring inter-site transferability of crop varieties in order to assess success of varieties evaluated in test environments for their production in target environments, a test for crossover type genotype x environment interaction and estimate of crossover point, estimation of heritability from variety trials conducted in complete or incomplete blocks, estimation and testing of genotypic and phenotypic correlations in plant traits from variety trials, modelling unstructured heterogeneity in field trials, Bayesian analysis of block designs, statistical assessment of screening techniques for heat tolerance in wheat, a procedure for analysis of data from long-term rotation trials, estimation of time-trends to measure the scale of sustainability of a farming system. Spatial models have been exploited to enhance the precision in field trials in barley, wheat, lentil and chickpea and efficiency of variety selection and for species abundance in bio-diversity studies.

BSS has enhanced efficacy of statistical computing culture at the Center by providing training and developing a number of computing modules for various designs and statistical analysis. These modules are also in use by scientists in the NARSs. The statistical software packages in use in the center include: GENSTAT, CycDesignN, SPSS, Stata, AGROBASE, SAS, S-PLUS, SYSTAT, StatXact. The Center scientists have been using a number of specialized applications available as Open Source software such as R

packages. In addition various other scientific packages and scientific presentation software are used in the center and supported by BSS. In several cases, the statistical analysis is carried out by BSS, results are discussed with the scientists and electronically shared.

Accomplishments of the BSS since 1989 are included in separate documents given in the links ([Annexure 1: Links to BSS Activities](#))

### **3. The Challenges**

To meet the needs of population and environment, today and in future, development of agricultural technologies and resource base will have to continue at a sustainable rate. Biometrical and statistical innovations will be needed to support the dynamic development. The R4D would require a) search and maintenance of diverse crop germplasm, b) evaluation of crop germplasm, c) cropping systems, likely to comprise of diverse crops, shrubs, grasses and trees, livestock and fish, with a conservation agriculture and crop rotations framework, and d) management of natural resources such as land, water and ecosystem, and e) tools to mitigate the risks due to adverse climates through technology and crop actuary. The statistical tools and techniques will be needed for efficient and reliable data collection and sound statistical tools to analyse data and draw inferences. In field trials for crop improvement new designs are needed to maximize genetic gain, accounting spatial variability and between plot competitions. While designing experiment for Sustainable Cropping Systems – harvesting synergies along time and space must be brought in and the focus should be to diversity system components in time and space including shared sole and intercropping in rotation. In case of resources management trials, the landscape parameters and competition between the species, e.g. shrubs and trees, must be factored in. Methods of statistical analysis should have effectively fit model to the data with various assumptions being examined and cross validation done.

In continued search or updates for better options on agricultural technologies to keep pace with the changing environment, scientists collect volumes of data or evidence, which remain largely ignored in evaluating the current datasets/evidences. Bayesian methods have the potential to incorporate prior information collected over a span of experimentation, in the areas of crop improvement and cropping systems research. ICARDA has already initiated application of the Bayesian approach in some situations, we need to examine its usefulness in all the other situations as well. Further details are specified in Section 5.

### **4. Current Dryland Agricultural Research Environments and ICARDA Research Platforms**

The recent unfortunate emergence of conflict in Syria, led the Center to decentralize and relocate its research activities coordinated and carried out at three Research Platforms in Morocco, Ethiopia and India, and specific theme research sites in Egypt, Lebanon, Turkey and Uzbekistan. The biometrics support to research scientists and their NARS partners are provided using the ICT tools and during visits to these locations, while the

development of techniques and tools are carried out by the BSS staff from their duty station.

With the reorganization of CGIAR research, the BSS support has also reoriented towards the activities of various CRPs, in particular the following CRPs where ICARDA is involved either as a partner center:

1. CRP DCLAS [Dryland Cereals and Legumes Agri-Food Systems]
2. CRP Wheat [WHEAT- Global Alliance for Improving Food Security and the Livelihoods of the Resource-poor in the Developing World]
3. CRP Livestock Agri-Food Systems [Livestock Agri-Food Systems: More meat, milk and eggs by and for the poor]
4. CRP CCAFS (Climate Change, Agriculture and Food Security)
5. CRP Genebanks
6. CRP PIM [Policies, Institutions, and Markets to Strengthen Food Security and Incomes for the Rural Poor]
7. CRP NH (Nutrition and Health) [Agriculture for Improved Nutrition and Health]
8. CRP WLSE (Water, Land, Soils and Ecosystems) [Improved natural resources management for food security and livelihoods]
9. CGIAR Big Data Analytics Platform

#### **5. The statistical issues that need to be addressed in above research environments**

The contexts of statistical issues underlying the challenges in Section 3 are described in the following in relation to the activities of specific research programs and CRPs:

#### **Biodiversity and Integrated Germplasm Management (BIGM)**

CRPs: DCLAS, Wheat, Genebank, CCAFS, NH + CGIAR Big Data Analytics Platform

Strengthening of the crop improvement program will require tools and techniques for:

- Efficient experimental designs and sound statistical analysis of data to enhance the efficiency of each crop germplasm development.
- Rationalization of the evaluation environment for a specific crop, identify lines for broad adaptation, identify lines for specific adaptation, identify stable locations, and quantify the rate of breeding progress and breeding success. Carry out GxE interaction study incorporating spatial variability and covariances between the environments

#### **5.1 Experimental designs**

Generally experimental designs used are in one-way incomplete blocks (alpha designs, augmented designs) to evaluate genotypes (un-replicated/ replicated). These designs are derived based on an optimization criterion which minimizes the average variance of pairwise genotypes/varieties comparison/contrast. In an ongoing crop improvement programs the genotype base keeps varying due to selection, inclusion and avoidance of desired genotypes, in which case genotype effects are justified as random. Therefore the designs are needed to minimize the average variance for predicted differences. This criterion thus also leads to maximization of heritability and genetic gain due to selection. Such designs should also take into account spatial variability in the field (Cullis, et al.,

2006) and degree of genotypic variation varying with the generation of material or with the crosses.

## **5.2 Statistical Analysis of field trials**

In the evaluation of genotypic response at a single or multi-environments and genetic gain due to selection (where applicable by the material type), need to incorporate kinship matrix and heterogeneity of error variance and genotypic correlation between environments. Unstructured heterogeneity of Error variation may arise in the field due to lack of any patterns in residual effects, and needs be accounted by the groups of plots or even group of genotypes; Other issues may be that heterogeneity in plot-error variances may be due to the groups of plots or with groups of genotypes, where clustering approaches – hierarchical, non-hierarchical – with a control of Type I error could be helpful.

Bring geo-reference even at a micro-scale in statistical analyses may improve the precision of the genotypes performance. *Spatial variability* masks the responses in the field trials. Challenge is to develop appropriate biometric techniques to improve quality of inferences by incorporating spatial information in all the field studies. This includes procuring and developing suitable methods, incorporating spatial variability patterns, for selection of high yielding and stable genotypes for specific adaptation, exploitation of genotype  $\times$  environment (GE) interaction; inheritance studies; bio-diversity studies; spatial modeling of climatic and soil parameters; models for yield responses from crops in rotation systems; models for agro-ecological characterization and clustering of cereal and legume growing environments, clustering of environments for disease reactions, prediction of extreme weather conditions, assessment of land degradation and its influential variables/ parameters; genotypic differentiation using hyper-spectral information on green cover of the crop genotypes at various points in crop growth duration (Singh, 2014).

Statistical techniques and computing tools will be needed to rationalize the evaluation environments representing the TPE (target population of test environments) in terms of clustering approaches, models for covariances between the environments, inter-site transferability and methods for exploiting GE interaction (see [Annexures 2 and 3](#) on CRPs DCLAS and Wheat pre-proposals).

*Examining the nature of GxE interaction using climatic information:* With advances in ICT (Information and Communication Technology), GPS (Geographical Positioning Systems), remote sensing satellites technologies, collection of affordable space-time reference data on temperature, rainfall, soil moisture at an experimental plot level as well as the crop growth parameters with time is a reality. Improved models for genotypic response and selecting the crop genotypes for specific adaptation to regions of various climatic conditions are needed in the light of these available variables.

*Stay ever-prepared for Climate Change:* The issues are to evaluate the effect of climate change on rationalization of the evaluation environment for each of the component dryland cereals and grain legumes; to model statistical distribution of extreme weather

events and estimate their return levels and return period/probability, and to predict the component crop responses under the extreme event scenarios. These aspects may be kept updated on a regular basis.

*Integrated Pest and Disease Management:* In the IPDM (Integrated Pest and Disease Management) areas, screening of desired genotypes (resistant to applicable insects or diseases) is carried out using scores which are in general ordered categorical or ordinal variables. The challenge is to include probability of misclassification of the scores, possibly under a Bayesian framework.

*Analysis of genotypic response data with Genetic markers at molecular level:* With the establishment of the biotechnology facilities, ICARDA produces a large volume of molecular marker data in addition to already on-going phenotypic data in relation to the environments. Techniques and tools are needed to select desirable crop genotypes with high yields or the breeding values and stable for various unfavorable stress factors including drought, cold, disease and pests. With the availability of genome-wide dense markers, MAS (marker assisted selection) and GS (genomic selection) techniques seek to build robust regression models to predict the breeding value of the individuals. With very large numbers of observed variables (markers) relative to the number of genotypes with or without kinship information available, development of multi-trait model is indeed a challenge (Meuwissen et al. 2001; Crossa et al. 2010). We also need to enhance the bio-informatics facility to store, manipulate, retrieve, register, analyze and interpret the marker data. An approach integrating molecular, phenotypic and environmental information would be required. This is the area where we need to regularly update our familiarity of the techniques and tools available with other institutions, e.g. ICRISAT, free software on the Internet, and use them to support the BIGM research.

### **Plant Biodiversity**

Statistical issues in this area focus on assessment of plant bio-diversity and selection of a diverse group or cluster of germplasm accession, normally based on the two approaches: 1) Genetic diversity in a population and sampling strategy for a mini-core collection, where an efficient sampling scheme is needed along with the determination of sample sizes and number of loci or quantitative traits for diversity assessment and 2) FIGS (Focussed Identification of Germplasm Strategy) where use of geo-reference is carried out in building model relating environmental variables with the traits of interest, including developing effective indices of weather variables e.g., principal components.

### **Seed Production**

The sampling design for seed quality surveys, study of shrubs diversity and effect of management practices by the community of small ruminant growers, and analysis of data from the seed quality surveys are needed. In the experimentation with mechanical harvesting experiments, we need to also incorporate of spatial variability in the models for a more precise evaluation of the treatments.

### **CRP Dryland system**

Mixed methodologies are for application to integrate quantitative and qualitative data in resilience activities with integrated technical and policy innovations for rangeland system; to rationalize/characterize barley-sheep systems will be applied for site characterization by grouping/rationalization and clustering; to develop total factor productivity indices; to participate at a statistical conference on Ecosystems, Environment and Sustainable Development; and techniques and tools are needed for valuation of Natural Resources.

### **CRP WLE**

For this Natural Resources Management theme, we need to carry out: analysis of data from water management and agronomic input trials, develop crop models incorporating the spatial variability patterns in the field layout.

### **DSIPS**

*Conservation Agriculture:* Design for CA is to evaluate genetic and agronomic input factors, some of which may require plots of relatively larger sizes (main plots) while the other factors require plots of relatively smaller sizes.

*Interaction between crops/grasses and shrubs/trees:* This area requires statistical criterion to develop optimal designs suitable to model interactions and competition between the species of crops, grasses, shrubs and trees, with due consideration to the landscape factors and the associated analysis of data.

*Crop rotations trials:* System specific design are needed to keep a long-term investment view with likely changes in the treatments over time, with crop rotation; Analysis of data will include modelling tempo-spatial variation, use of climatic variables and planting date, accounting for plot- error covariance, for an evaluation of productivity and sustainability including the time-trend, time to terminate the trial. Preservation of data is highly valuable in a structured DB format to save resources by avoiding repetition, add value in future exploration and a readily available resource or environment for training.

### **SEPR**

*Crop Actuary:* Actuaries use the probabilistic, mathematical and financial theory and tools to predict risk of failures and develop insurance policies in health and industry sectors and can be evaluated in agricultural systems to save the farmers from life risking decisions arising from the unpredictability of rainfall, drought frequency and frost severity.

### **Cross-cutting**

#### **5.3 Bayesian approach**

Both the frequentist and Bayesian paradigms of statistical inference may be continued for data analysis. There is however a need to develop user friendly steps to incorporate prior information in the Bayesian framework, and compare the performance in each specific research area.



## **5.4 Statistical Computing**

Statistical packages such as Genstat software and other affordable ones, such as CycDesigN, ASREML, SAS, SPSS, Stata, etc. may be continued for general purposes of designing and analyzing the data on with inter-disciplinary research. Scientists may be encouraged to use the software of their competencies. We will continue to develop ICARDA online-biocomputing modules using Genstat and extend to use R-codes and its specialized packages to keep pace with the developments outside ICARDA (see [Annexures 4 and 5: DRAFT on BioComputing II](#))

*Specialized Applications software:* Depending on the requirement of a specialized design and analysis situation, BSS team will continue to routinely download, install, learn and provide support on the use of new software as an when they would be needed.

## **6. Experimental Databases:**

### **6.1 Raw data**

Raw (experimental-unit-wise, e.g., a field plot-wise) data from experiments are ever invaluable for not only a short-term exploitation and use in developing scientific information and knowledge product, but add value even on a long-term basis when one evaluates a series of data (e.g., multi-environment trials, MET) and prior information available. For the breeding activities, BIGM has stored the field plot data sets using a commonly used platform ICIS and its updated versions and related products ICIS and Breeding Data Management, with/without recently developed an add-on “F1 crosses” for generating the crossing blocks.

Datasets from molecular marker studies and genomics have specialized data systems and tools to retrieve and analyse. A structured database system(s), such as ICIS, at the Center will have the advantage of sharing CRP wide standard protocol and may be regularly kept updated for any changed pedigree naming convention and or ICT-environment in future. With a standard agreed for contributing data to a central system, steps for ensuring (high) quality data becomes essential and will result in enhanced quality of research outputs. Such a system will be extremely helpful to individual scientists as well as to the succeeding scientists.

Datasets from natural resources areas including long-term rotations in Syria, agronomic and water resources trials are available on individual scientist computers. These datasets needs to be organized in structured databases for a more efficient utilization of the investment and generation of science product.

### **6.2 ICARDA Data Management Open Access Policy:**

Refer to “CGIAR Open Access and Data Management Plans & Implementation (Article 4.1.9) which states “Open Access and Data Management Plans should be prepared in order to ensure implementation of this Policy. Such Plans shall, in particular, outline a strategy for maximizing opportunities to make information products Open Access”.

BSS will continue to contribute develop a research data management strategy for all research programs in the implementation of research quality management systems to

ensure that all the research activities are producing *high quality data* and to the Strategic Results Framework (SRF) of the CGIAR. BSS's immediate role would be to advise the participating scientists on the sound design of experiments and surveys to enhance the research quality and the on consistency checks of the data to enhance the data quality.

Presently, Geoinformatics Unit (GU) at Amman has implemented use of a HP StoreServ 7400 solution with total 350TB Storage Capacity. (<http://geoagro.icarda.org/>). The steps followed for CRP DS on data curation, management and sharing would be helpful for the other programs well.

## **7. Training and Capacity Building**

NARS scientists develop agricultural technologies with or without collaboration with ICARDA scientists. National Coordination Meetings have expressed need of ICARDA involvement to enhance the quality of science in NARSs, which can be achieved by sharing knowledge on statistical concepts and methodologies for designing field experiments, and biometrical and statistical techniques applied in agricultural research and use of statistical software. A training course also provides a workshop opportunity to carry out statistical analysis of their data and prepare a draft research manuscript for publication. These activities can be carried out as short-term group courses and individual degree or non-degree training at various locations. The training courses have been undertaken in a range of diverse areas of interest (see Annexure 6: Commonly used topics in the training courses). We plan to keep the training extended to cover some new areas involving breeding value estimation using molecular marker data in collaboration with BIGM, Bayesian analysis of field trials, tempo-spatial modelling of trends in land degradation and other indicators of climate change using geo-informatics and remote sensing data in collaboration with GU.

## **8. Organization of the BSS Activities**

### **8.1 Activities**

To support the current research programs/units:

BIGM: Biodiversity and Integrated Gene Management Program

IWLM: Integrated Water and Land Management Program

DSIPS: Diversification and Sustainable Intensification of Production Systems Program & SEPR: Social, Economic and Policy Research Program

GISU: Social, Economic and Policy Research Program,

BSS activities have been classified with following titles and objectives:

1. BSS-1: Advisory Support to Research Projects
  - To render biometrical and statistical consultancies, and assistance in planning and analysis and drawing statistical inference from the experiments/surveys conducted under various projects at the Center and the collaborative projects in the NARS; to review the scientific documents.
2. BSS-2: Exploitation of Advanced Biometrical and Statistical Techniques
  - Develop statistical methods to fill the gaps in biometrical and statistical techniques in designing experiments and surveys, and modelling of data

3. BSS-3: International Crop Information System (ICIS) Support
  - To provide support for maintaining the ICIS and carry out development of interfaces
4. BSS-4: Bioinformatics Support
  - To provide bioinformatics support for developing Genetic Resources, Crop and Genomic Information and Analysis Systems at ICARDA
5. BSS-5: Biometrical and Statistical Computing
  - Develop biometrical, statistical and bioinformatics computing modules, document them and make them accessible on-line within the Center and to the NARSs.
  - Implementation of Bio-Computing Online, Phase II
6. BSS-6: Supports on Statistical Software
  - To provide installation, maintenance and user support on all statistical packages that are supported by BSS and assistance in data analysis
7. BSS-7: Training
  - Training in biometrical and statistical methods and associated computing
8. BSS-8: Support to DDG-R office
  - To assist DDG-R on research management related matters

## **8.2 Hardware and software:**

- To position the BSS to assist researchers in all the research projects to meet their biometrics and bioinformatics needs, including methodology and software, and training of in-house and NARS researchers in the BSS areas, we need to:
  - Procure the computer hardware, software and analytical tools for storage, update, retrieval; GU server provides a storage media at Amman. Needs to meet the requirement of other locations.
  - Procuring necessary statistical and other specialized software on design of experiments and surveys, and analysis of data

## **8.3 Personnel**

Existing staff are:

- M. Singh, Senior Biometrician
- K. El-Shamaa, Specialist-Experimental Research Informatics

## **9. List of ongoing BSS activities Workplan for 2015**

See [Annexures 7](#). The BSS activities for BIGM are broadly characterized in [Annexure 8](#).

## **10. Networking**

- In-house community of practice
- Statistical software email distribution list
- CG Biometrician's list

## REFERENCES:

- CGIAR Big Data and Analytics Platform  
CRP DCLAS preproposal  
CRP Wheat pre-proposal
- Crossa J, Campos G, de L, Pe´rez P, et al. 2010. Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. *Genetics* 186: 713–724.
- Cullis, B., Smith, A. & Coombes, N. (2006). On the design of early generation variety trials with correlated data. *Journal of Agricultural, Biological and Environmental Statistics* 11: 381-393.
- Singh, M. (2014). Statistical Research Issues in Crop and Cropping System Experiments for Enhancing Food Security Support. *Journal of Indian Society of Agricultural Statistics*, 68(1):1-10
- Meuwissen T.H.E., Hayes B.J., Goddard M.E. (2001) Prediction of total genetic value using genome wide dense marker maps. *Genetics*, 157, 1819–1829.

## Annexures

### Annexure 1: Links of BSS documents

1. [2014 Biometrics and Statistics Harvest](#)
2. [2013 Biometrics and Statistics Harvest](#)
3. [2012 Biometrics and Statistics Harvest](#)
4. [2011 Biometrics and Statistics Harvest](#)
5. [Period 2008-2010](#) and [a consultancy report 2010](#)
6. [Period 1989 – 2007 \(Overview of activities\)](#)

### Annexure 2: Extracted from CRP DCLAS Pre-proposal for 2<sup>nd</sup> Call (2015), pages 41-42

#### **CoA 3.1: ENVIRONMENTAL CLASSIFICATION, TPEs AND PHENOTYPING:**

The prediction of genotype performance for yield and other agronomic traits in a target population of environments (TPE) is made possible by the correlation between genotype performance and environments using data on relative performance of genotypes in multi-environment trials (METs) (Basford et al., 2002).

- For the combined realization of success and efficiency gains in a complex crop improvement program involving twelve target crops grown in at least three different production environments (arid, semiarid, and humid tropic), it is important to (1) identify and understand the multiple production systems and TPEs of the different target crops of the program, (2) collect, update and classify environmental characterization data of both the test and the target population of production

environments, and (3) utilize spillover opportunities for both testing and release across locations, production systems, and regions.

To this end, during 2017-2022, we aim to:

1. systematically catalogue, characterize and classify all production environments of the target crops in the target countries,
  2. identify and define separate or overlapping TPEs for the target crops and their target production environments,
  3. determine the existence, or lack, of representative selection and testing locations
  4. complete spatial characterization of existing representative selection and testing locations, and establish new collaborative location networks (with CGIAR or non-CGIAR partners as required) where gaps exist
  5. establish at least one pilot-mode managed-stress testing location to enable breeding for one critical trait, drought tolerance
  6. systematically apply new and more precise statistical methods of experimentation that combines design information with spatial adjustment within and between trials
  7. identify controlled-environment phenotyping parameters (physiological or metabolic) predictive of field performance, improve existing models that address GxExM, and design target plant ideotypes
  8. explore in pilot mode the feasibility and practical relevance of remote-sensing for field phenotyping in our target crops in the target ecologies, and
  9. establish centralized biometrics facilities, cloud-based data warehousing and computing, and pedigree management and tracking systems for the application of optimal statistical design, analytical protocols and appropriate metadata standards
- The recent convening of the partners involved in the IAVAO Partnership Platform (PP) in Montpellier identified the development of a field experimental network as a priority objective and a key landmark towards the implementation of this integrative approach. This network will aim at providing experimental and breeding sites representative of the TPE in West Africa and equipping them to multi-environmental trials for GxExM analysis, genetic-material validation, and phenotyping of adaptive traits in West Africa.

Annexure 3: Extracted from CRP Wheat Pre-proposal page 23-24.

**Biometrics research to develop models that enhance efficiency of breeding programs.**

- to enhance efficiency and genetic gains in breeding.
- Development of models that accurately predict performance under multiple environmental conditions by using covariates such as maximum and minimum temperature, solar radiation, and precipitation will reduce the need for costly and often difficult advanced testing.
- Strategic interventions, issues and challenges include: a) Develop models for genomics-enabled prediction for diverse breeding populations, including gene bank accessions and elite populations; b) develop predictive models that incorporate

genotype by environment interaction (GxE), including weather data; c) increase the accuracy of predictive models by combining GxE with high density molecular markers and pedigree information; d) develop statistical genomic models that incorporate non-normally distributed, categorical response traits (e.g. disease symptom scores); e) improve the precision of GS models by using high-fidelity, high-throughput phenotypic data reducing the time and cost of phenotyping; and f) develop selection indices that incorporate the diverse types of data (e.g. digitized image, weather and genomic data).

#### **Develop and validate novel genomic selection (GS) approaches for wheat breeding.**

GS is a relatively new breeding technology based on using genome-wide marker data and phenotypic data to predict breeding values for quantitative traits that have not been extensively phenotyped (Meuwissen et al. 2001).

- Several prediction models will be investigated, including: a) genotype-by-environment interaction (GxE) models incorporating genome-wide markers and pedigrees; b) multivariate models that combine either genomic or pedigree relationship estimates and data on several traits including high-throughput yield proxy traits suited for HTPTS (see below) and metabolite levels from laboratory assays; and c) new relationship matrices models, including those generated using data from images and sensors.

#### **Develop and validate novel high-throughput proxy trait selection (HTPTS) for wheat breeding.**

In contrast to GS, HTPTS is a new prediction-based selection approach being pioneered with WHEAT colleagues at Cornell and Kansas State University. HTPTS allows rapid and inexpensive data collection throughout the crop cycle for multiple traits correlated with yield, e.g. canopy temperature and normalized difference vegetation index (NDVI) using aerial remote surveillance vehicles.

- HTPTS will be evaluated for yield under optimal conditions using small plots for large numbers of selection candidates, for which genome-wide genotyping is not currently feasible.

#### **Develop and validate multi-trait prediction models and selection indices for wheat breeding.**

As predictions are increasingly used to enable selection based on multiple traits, selection indices become increasingly important because breeding value predictions from GS and HTPTS are statistically biased and cannot be interpreted as actual trait values.

- Multi-trait prediction models produce predictions that can be weighted to generate an optimal selection index to maximize gain for total genetic value (including all traits) (Mrode 2014).
- Multi-trait prediction models and strategies for their implementation will be evaluated via simulation studies to recommend their use in wheat breeding programs.
- The improvement in efficiency of selection for multiple traits will be assessed using simulation studies.

#### Annexure 4: List of Online Biocomputing module

##### **Experimental designs modules (8 modules)**

**AD-** Augmented Designs in one-way blocks  
**ISD-** Initial Screening Design for unreplicated test-entries with repeated check rows  
**RCBD-** Generates Randomized Complete Block Designs  
**SqLat-** Generates randomized plans from Square Lattice designs (one-way blocks)  
**D2FCB-** Generates randomized plans for two treatment-factors in complete blocks.  
**D3FCB-** Generates randomized plans for three treatment-factors in complete blocks.  
**DSPlot-** Generates randomized plans for standard split-plot experiments.  
**DSPSPlot-** Generates randomized plans for standard split-split-plot experiments.

##### **Standard Analysis (10 modules)**

**CRD --** analyses similar trials in completely randomized designs;  
**RCBD --** analyses similar trials in RCBD with a Single Treatment-Factor  
**AUG1Way-** in augmented designs with 1-way blocking system  
**URNB-** in an Un-Replicated design with No Blocking  
**A2FCB-** with two treatment-factors in complete blocks.  
**A3FCB-** with three treatment-factors in complete blocks.  
**ASpPlot-** in standard split-plot experiments.  
**SpSpPlot-** in standard split-split-plot experiments.  
**IBD-** in Incomplete Block design with a single treatment-factor  
**IBDHG-** in IBD with estimates of heritability within each group of genotypes

##### **Quantitative Genetics (9 modules)**

**GCCB-** Genotypic, phenotypic and environmental Correlations – RCB  
**GCIB-** Genotypic, phenotypic and environmental Correlations – IBD  
**PA-** Genetic Path Analysis using data on genotypes means  
**LxT-** Line x Tester analysis of data from similar experiments in RCB  
**Selection Index CB-** Develops an index for selection – RCB  
**Selection Index IB-** Develops an index for selection – IBD  
**Fit 2Ratios, Fit 3Ratios, Fit 4Ratios-** Fits given ratios

##### **Multi-environment trials (7 modules)**

**MET CB-** Analysis of Multi-Environment variety Trials conducted in RCB  
**MET IB-** Analysis of Multi-Environment variety Trials conducted in IBD  
**Stability4 CB-** Computation of stability Type 4 index –RCB  
**Stability4 IB-** Computation of stability Type 4 index – IBD  
**MET Means-** Analysis of Multi-Environment variety Trials using Means and SEs  
**Stability4 Means-** stability Type 4 index from multi-location multi- year..  
**Inter-site Transfer Means-** inter-site transferability of crop varieties

##### **Spatial Analysis (6 modules)**

**SPCB-** Spatial analysis of data from a Complete Block design.  
**SPIB-** SPatial analysis of variety trial data from an Incomplete Block design.  
**SPUR-** SPatial analysis of variety trial data from an Un-Replicated design  
**SPCBHG/SPIBHG/SPURHG ...**with est. heritability within each group of genotypes

## Annexure 5: Bio-Computing Online, Phase II

**Bio-Computing Online, Phase II: a new generation of data science**  
*(Draft Project Proposal by Khaled El Shamaa)*

### **1. Executive Summary**

Importance of web technologies increases at ICARDA because of decentralization strategy for research, where needs to access computing and biometrics resources remotely become essential. In addition, this kind of services provides an opportunity to deliver high-quality international public goods to NARs. On the other hand, it will be in line of the CGIAR support open access and data management policy. The potential short-term application is to develop statistical services tool for the CGIAR Research Program on dry land systems monitoring and evaluation website (i.e. CRP-DS M&E <http://drylandsystems.cgiar.org/mel/index/home/>).

### **2. Statement of Problem: the “Why?”**

- The need to access biometrics resources remotely because of ICARDA decentralization strategy.
- The need to deliver public goods accessible for NARs and recognizable by donors.
- Help in meet the requirements of open data and open access trend in the CGIAR consortium.
- Lessons learned from our experience in running the current Bio-Computing online system since 2003.
- Cut down the cost of the statistical packages licensing specially for temporary/seasonal users.
- Free open source R language capacity building trend.

### **3. Objectives: the “What?”**

We propose to develop existing Bio-Computing online system for customized analysis procedures (called modules) where end-users can upload their data directly to the website in popular Excel format, and use online interactive interface to define and set all related analysis parameters and arguments. Analysis will perform at the backend server using R language. We will use Shiny framework technology to build the web interfaces. Outputs will present directly at the end user browser as well as deliver in a proper way to download.

### **4. What is different from the current system?**

- Backend analysis will process using open source R language instead of proprietary GenStat



application, this follow community trend and will have no hidden license cost or any legal issues.

- Front end user interface will use friendly web page instead of email system in the current system, this should need less instructions, and will reduce the ratio of failures in formatting requests as required.
- Response will be automatic instead of current half-automatic system, in other words we will need less time to monitor system usage instead of standby operator to process requests.
- System can present as public good in more straight forward way instead of circulating current system in CD's and introduce in training courses.

## **5. Technical Approach: the “How?”**

5.1. The overall design process steps for this proposed system in general terms are:

- 5.1.1. Setup the hosting server and infrastructure (at <https://www.shinyapps.io>).
- 5.1.2. Select the modules to implement (e.g. t-Test, correlation, and PCA).
- 5.1.3. Develop R script to perform selected analysis.
- 5.1.4. Test the developed R script and verify output results.
- 5.1.5. Deploy verified R script in Shiny server to deliver it as a web based application.
- 5.1.6. Publish it as public good to our NARs.

5.2. Additional approaches may include:

- 5.2.1. Distribute our code as R package (Pros: we can publish it at [CRAN](#), the official R packages repository like what CIP did for [Agricolae](#) package; Cons: end users should have well programming experience in R language).
- 5.2.2. Install and setup our [RStudio Server](#) and configure our own environment (Pros: no need to setup the R language environment at the end user desktop, and can be used to access high performance computing resources remotely. Cons: still end users needs well programming experience in R language, also server requires extra cost for maintenance in term of time and money).
- 5.2.3. Deploy our R analytics script using [DeployR](#) integration technology (Pros: it provides an API to deploy our R analytics inside web, desktop, and mobile applications, Cons: it is more like tool for applications developers rather than end user public good).

## **6. Proof of the concept:**

I developed a demo application to analyses data of an experiment conducted in completely randomized design (CRD) to evaluate a single treatment factor. This program uses analysis of variance (ANOVA) procedure to compute main effects of the factor, table of means, standard error of means (SE), standard error of differences of means (SED), least significant differences of means (LSD), and coefficients of variation (CV%). Output enhanced also by present residual plots and interactive mean bars.

One can try online demo by click the following link: <https://icarda.shinyapps.io/OneWay> and check the source code of this application here: <https://github.com/icarda/OneWay>

## **7. Which modules we are going to start with?**

The selection criteria focus on straight forward analysis with few options and minimum data processing, we will start with the following cases:

- t-Test (paired/un-paired, alternative hypotheses, variances equality, normality assumption, box-plot)
- Correlation test (process bunch of columns, test of significance, all pairs scatter plot)
- Principal Components Analysis (scaling, matrix of variable loading, importance of components, biplot)

and cover the other modules in due course of time.

## **8. Deliverables:**

- Number of available modules as public goods.
- Statistics of usage by end users.

## **9. Project Management: “How and When?”**

### **9.1. Setup the Hosting Server**

In this task and for each round of annual working plan, developer (i.e. K. El-Shamaa) should search for best hosting offer, apply required purchase request, and collaborate with responsible PSD and Finance staff who will process this order. This task needs one week on yearly base, and outcome is up and running Shiny hosting server with proper management tools.

### **9.2. Select Modules to Implement**

In this task and for each round of annual working plan, biometrician and developer (i.e. M. Singh and K. El-Shamaa) will agree on the list of analysis modules to add the system during the year based on the technical complexity and ICARDA scientists needs. This task may need one week on each planning round, and outcome would be the list of modules to add with related references.

### **9.3. Develop R Script**

In this task and for each new module, biometrician (i.e. M. Singh) will provide the set of sample input files associated with desirable analysis output to developer (i.e. K. El-Shamaa) who will develop standalone R script to perform required analysis. This task needs two weeks or longer for each new module, and outcome will be a ‘ready to execute R script’ that, on processing the sample input files, will generate required output files.

### **9.4. Test R Script**

In this task and for each new module, developer (i.e. K. El-Shamaa) will provide related R codes developed in the previous step including example input and output files to the biometrician (i.e. M. Singh) who will test and verify results by cross comparing to other certified statistical applications. Developer (i.e. K. El-Shamaa) will participate in this activity by debugging related R code to fix any reported issue. This task may need one week for each new module, and outcome will be a ‘verified R script’ ready to deploy online.

#### 9.5. Deploy on Shiny Server

In this task and for each new module, developer (i.e. K. El-Shamaa) will work on the verified R code that is tested in the previous step to develop web interface required for deploy it on Shiny server (i.e. develop both of server.R and ui.R scripts) and upload it to the hosting Shiny server. This task may need two weeks for each new module, and outcome is up and running online version of related analysis script.

#### 9.6. Test Online Implementation

In this task and for each new module, developer (i.e. K. El-Shamaa) will send link address of the ready to publish online web application that implements related module analysis to the biometrician (i.e. M. Singh) who will test it online and send his feedback to refine user interface to be clear and user friendly as much as possible. Notices and advices will send back to the developer (i.e. K. El-Shamaa) to carry out what is technically visible. This task needs one week for each new module, and outcome is bio-computing web application ready to publish as public good.

#### 9.7. Publish it as Public Good to NARs

In this task and for each module, biometrician (i.e. M. Singh) will send the final link address of the stable module web application associated with a paragraph describe it to CODIS focal point who is responsible to publish ICARDA public goods at our official website as well as proper social media channels. This task needs one week for each new module, and outcome is a new item added to the list of available modules at ICARDA Bio-Computing online service.

### **10. Budget: “How Much?”**

“Free” hosting plan from RStudio vendor at shinyapps.io enable us to deploy our proof of concept applications to the cloud for FREE. But it is limited to host up to 5 applications only, and provides 25 active hours (i.e. the hours our applications are not idle), also RStudio branding and the RStudio logo will appear on all applications. While “Basic” hosting plan has no restriction on the number of hosted applications, active hours are increased to 250, no RStudio branding will appear on our applications, and we will get an email support. This will cost us around \$500 USD per year, as per current estimate. For more information: <http://www.rstudio.com/pricing/#ShinyApp>. On the other hand, the plan assumes that we need 20 person-weeks in 2015-16 to setup the platform, earn required knowledge and experience, and implement the suggested three modules in this proposal (i.e. t-Test, correlation, and PCA analysis). Later on, adding new module to the system will need in average six person-weeks.

### **11. Team Qualifications: the “Who?”**

- Leader: Khaled El-Shamaa (main system developer, R programmer)
- Participants: Murari Singh (review statistical algorithms and verify analysis results), any ICARDA scientist who has R code or algorithm to implement online (e.g. Abdallah Bari and Miguel Sanchez-Garcia)

### ***Appendix – Related R Technologies Glossary***

**R** is a language and environment for statistical computing and graphics. It is a GNU project, which is similar to the S language and provides a wide variety of statistical (linear and nonlinear modelling, classical statistical tests, time-series analysis, classification, clustering ...) and graphical techniques, and is highly extensible. The S language is often the vehicle of choice for research in statistical methodology, and R provides an Open Source route to participation in that activity.

<http://www.r-project.org/>

**RStudio** is an integrated development environment (IDE) for R. It includes a console, syntax-highlighting editor that supports direct code execution, and tools for plotting, history, debugging, and workspace management.

<http://www.rstudio.com/products/rstudio/features/>

**Shiny**, an easy web applications in R: It is an open source R package that provides an elegant and powerful web framework for building web applications using R. Shiny helps you turn your analyses into interactive web applications without requiring HTML, CSS, or JavaScript knowledge.

<http://www.rstudio.com/products/shiny/>

**RStudio Server** lets you access RStudio from anywhere using a web browser. RStudio Server Pro delivers the security, centralized resource management, metrics, and commercial support professional data science teams need to develop in R at scale.

<https://support.rstudio.com/hc/en-us/articles/200552306-Getting-Started>

**DeployR** is an integration technology for deploying R analytics inside Web, desktop, mobile, dashboard applications and backend systems. DeployR makes it easy for application developers to collaborate with R programmers and integrate R analytics without any R programming knowledge. Using analytics Web services, it solves an integration problem faced by those adopting R-based analytics alongside existing IT infrastructure.

<http://deployr.revolutionanalytics.com/documents/about/>

**High-Performance and Parallel Computing with R:** This CRAN task view contains a list of packages, grouped by topic, that are useful for high-performance computing (HPC) with R. They are defining 'high-performance computing' rather loosely as just about anything related to pushing R a little further: using compiled code, parallel computing (in both explicit and implicit modes), working with large objects as well as profiling.

<http://cran.r-project.org/web/views/HighPerformanceComputing.html>

**Revolution R Open (RRO)** is the enhanced distribution of R from Revolution Analytics. It is a complete open source platform for statistical analysis. It is based on (and 100% compatible with) the latest release of the R language, and includes additional capabilities for performance, reproducibility and platform support.

<http://mran.revolutionanalytics.com/documents/rro/open/>

## Annexure 6: Topics for the training courses

*Basic principles of experimental design, design and analysis of single factor experiments in randomized complete block design (RCBD), design and analysis of variety trials in*

*alpha-designs/lattices, analysis of multi-location variety trials conducted in RCBDs/incomplete block designs, stability analysis, spatial analysis of data, design and analysis of two-factor factorial experiments, split-plot experiments, strip-plot experiments, crop-rotation trial; Design and analysis of on-farm trials; Design and analysis of cross-over trials; Analysis of covariance; Elementary probability theory, statistical methods—standard distributions, estimation and testing of population means and proportions; Regression analysis; Sampling Techniques; Multi-variate analysis—MANOVA, Principal component analysis, principal coordinate analysis, discriminant analysis, correspondence analysis; Diversity indices; Quantitative genetics—variance components, heritability, genetic gain due to selection, genotypic correlation; Estimation of return level/period of extreme events.*

## Annexure 7: List of BSS Activities in 2015 (as of August 2015)

### **1. BSS-1: Advisory Support to Research Projects**

#### CRP DC

- Data analysis plan, Data management, Wrote Genstat codes for analysis (variance components, heritability, genetic gain, BLUPs) for individual trials. (RPS Verma)
- Draft manuscript on “Identifying barley genotypes for optimum input conditions in the NAWA Regions”, Verma et. al.
- Analysis of data, from augmented design, on agronomic and micro-nutrient traits to describe variability/diversity, correlation between traits, clustering of genotypes and ordering them using PCA. (S. Gyawali)
- Fe and Zn data analysis (S. Gyawali)

#### CRP DS

- Review of the CRP DS Workplan documents submitted for the flagships (WANA) and identify the list of survey activities to be undertaken for NAWA, WAS&DS, E&SA, SA and CA flagships.
- Review questionnaire “Baseline survey for CRP Dryland Systems for CA”
- Review the document “CGIAR Research Program on Dryland Systems, 2014 Annual Performance Report”, planned for submission to CGIAR Consortium Office, 10 March 2015.
- Review Survey questionnaires on activities, including gender contribution, on community structure, labor service providers, collectors, processors and producers, in Mahon and Sayaga action Sites in Burkina Faso in the WAS flagship.

#### WLE

- Analysis of data on olive land and water productivities on seed and oil yields from the trials in Syria and Morocco. Develop Genstat codes.
- A write-up on M&M and interpretation (Programme for the Development and Dissemination of Sustainable Irrigation Management in Olive Growing, CFC; Vinay Nangia)

#### CLAP (Afghanistan)

- Supervision on data checks on multi-locational replicated trials in RCBD on chickpea, lentil and mung bean.
- DRAFT manuscript development for studying GxE Interaction.
- Analysis of on-farm demonstration trials in chickpea.
- Update the DRAFT manuscript on “Enhancing food security by introducing improved variety of food legumes in Afghanistan”
- Analysis of on-farm demonstration trials in mung bean.
- Draft “Genotypic variability, yield response to the environment in Afghanistan and specific adaptation of food legumes  
(Yashpal Sahrawat)

#### Bilateral/CRP GL

- Analysis of lentil data on grain Fe and Zn of all 314 lentil genotypes, from 4 environment in India.
- Analysis of METs data for a NARS scientist

#### Egypt

- Analysis/estimation of heritability on mean basis for data in an alpha design. Genstat codes and outputs were emailed to her.

## **2. BSS-2: Exploitation of Advanced Biometrical and Statistical Techniques**

Develop methods, compute, prepare a draft manuscript or revise an existing one, reanalysis, response to reviewer's:

- Adnan Al-Yassin, Murari Singh and Michael Baum (2015). Genetic variation in winter barley and selection of high yielding lines. DRAFT.
- DRAFT: Patil, P., Dutta, D., Singh, M. and Biradar, C. (2015). Quantification of the terrestrial phytomass and carbon in the mountainous forest ecosystem using remote sensing and in-situ observations. An invited special session on “Forests, biodiversity and terrestrial ecosystems”, at the 36th International Symposium on Remote Sensing of Environment (ISRSE), 11-15 May 2015, Berlin, Germany.
- Chaubey, Y.P., Sarker, A. and Singh, M (2015). Power Transformations: An Application for Symmetrizing the Distribution of Sample Coefficient of Variation from Inverse Gaussian Populations. Accepted as a book chapter in Taylor & Francis publication.
- Patil, P., Atassi, L., Omary, J.E., Tulaymat, M.F., Singh, M., and Biradar, C. (2015). Mapping and monitoring of agricultural production systems for the better interventions of crop management. At an invited special session on “Agricultural ecosystem monitoring, assessment, conservation, and service sustainability”, The Fourth International Conference on Agro-Geoinformatics- 2015 - Istanbul, Turkey. DRAFT.
- Sarker, A. and Singh, M. (2015). “Improving breeding efficiency through application of appropriate experimental designs and analysis models: a case of lentil (*Lens culinaris* Medikus subsp. *culinaris*) yield trials” Field Crops Research. REVISION/REANALYSIS/
- Design and develop jackknife estimation of GC, GxEI, PC and SDC from RCB,

IBD, RCBMETs, IBDMETs.

- Develop an R program for comparing populations for total alleles arising from several markers, for Dr Udupa. New: In addition to the permutation test, also included normal approximation using square-root transformation on  $(x+3/8)$ . (It is based on the method in FU, YONG-BI(2010).
- Siraj et al (2015) Bayesian estimation of GxE interaction in Sorghum variety trials” submitted to Communications in Biometry and Crop Science. Revision/analysis
- DRAFT: A.A. Niane et al. (2015). Impact of grazing management...diversity, Syria'. Reanalysis plan, Revision of the manuscript in the light of comments from The Journal of Arid Environments.
- Fouad Maalouf, Miloudi Nachit, Michel Edmond Ghanem and Murari Singh (2015). Evaluation of faba bean breeding lines for spectral indices, yield traits and yield stability under diverse environments. Accepted in Crop and Pasture Science. Revision/Analysis
- Singh, M, A.B. Damania and Y.P. Chaubey Submitted a manuscript “Plant Genetic Diversity: Statistical methods for analyzing distribution and diversity of species”. Accepted as a book chapter. Re-write-up/analysis
- DRAFT: Gyawali et al. (2015). Genetic variability in agronomic traits and micronutrients in Kernel of wild and cultivated barley germplasm of ICARDA. For Euphytica. Write-up M&M.
- Omer, Siraj Osman, Abdel Wahab H Abdalla and Murari Singh (2015). Bayesian estimation of genotypic and phenotypic correlations from crop variety trials. Accepted in Crop Breeding and Applied Biotechnology. Revision/reanalysis
- Sarker, A, A.H. Rizvi, M. Singh and S.K. Agrawal (2015) Genetic variability for nutritional quality in Lentil (*Lens culinaris* Medikus Subsp. *culinaris*). For GRACE. Revision.
- DRAFT: A. Karadsheh et al. (2015). Genome Wide Association Mapping for Yield and Yield Components in Jordanian Barley (*Hordeum vulgare* L.) Landraces Grown under Drought-Prone Conditions. Write-up M&M, edits.

### **3. BSS-3: International Crop Information System (ICIS) Support**

- 

### **4. BSS-4: Bioinformatics Support**

- An R program for comparing populations for total alleles (it is a measure of diversity) arising from several markers and associated permutation test.

### **5. BSS-5: Biometrical and Statistical Computing**

- Initiate Implementation of Bio-Computing Online, Phase II

Customize:

- RCB module: for heritability on mean basis, genetic gains, line plots/plot-residuals, multiple comparisons
- IBD module: for heritability on mean basis, two more estimates, genetic gains, multiple comparisons, variance components

- METCB module: for heritability, mean basis, retain only average link.
- METIB module: for heritability, mean basis, two more estimates; retain only average link
- SPCB module: for two more criteria for model selection, heritability on mean-basis and genetic gain.

#### **6. BSS-6: Supports on Statistical Software**

- Support with Genstat, Agrobases, SPSS, Stata
- Renewal of Genstat licenses, CycDesignN, ASREML, SAS

#### **7. BSS-7: Training**

- Statistical Design, Data Management and Statistical Analysis in Agricultural Research in Afghanistan, 8-12 Mar 2015. 3 participants (1 woman) from Afghanistan
- Delivered a lecture “Design and Analysis of Water Resources Experiments” on 18 May in the training course “Improving Agricultural Water Productivity (with emphasis on irrigated production systems), 3 – 21 May 2015, ICARDA, Amman, Jordan (12 participants)
- Deliver a 1-day lecture on “Design and analysis of legume variety trials”, 21 May 2015 as a part of “Specialized Training Course on Food Legumes Improvement” from 17 – 26 May 2015, Tal Amara- Terbol, Lebanon (18 participants)
- Selected Advanced Statistical Techniques using R, June 2015, Innovation and Technology Transfer Center, Aleppo University, Aleppo, Syria (by Mr Khaled El Shamaa)
- Design and analysis of on-station and on-farm trials for lowland countries of the SARD-SC Wheat Project, 27 April – 5 May, 2015, Addis Ababa, Ethiopia
- Design and analysis of on-station and on-farm trials for highland countries of the SARD-SC Wheat Project, August 2015, Venue=??
- Statistical Design, Data Analysis and Biometrical Techniques in Agricultural Research , 1-12 November 2015, Amman, Jordan
- Basic Field Plot Techniques, Design of Experiments, Data Management and Statistical Analysis, 17-26 Nov 2015, New Delhi (Afghanistan participants).
- Advanced Biometrical Techniques in Crop Improvement”, 30 Nov - 9 Dec, 2015, New Delhi, India for participants from SA&C Regional Program
- 

#### Degree students

- Siraj Omer, Sudan for Ph. Degree. (Supervision, edits of thesis, manuscripts for publications, communication with SAS Administration)
- Miss Aline Abou Naoum, Lebanon for an M. Sc. degree with Dr. F. Maalouf. (Analysis of her experimental data)

#### **8. Support to DDG-R office**

- Matters as and when arise to assist DDG-R on research management related matters including:
  - Minutes of Annual Planning Meeting, 1-4 Feb 2015



- Tabulation of Programs x CRP Output/Workplan 2015
- Land required for experiments, and a draft email for DDG-R
- Decentralization issues on Equipment lists, bilateral funds tables, etc.
- Publication statistics

Annexure 8: BIGM – Crops barley, Chickpea, lentil, faba bean, bread wheat, durum wheat

**BUS title: Biometrics and Statistics Support to BIGM**

**Exploitation of Genotype x Environment Interaction** in barley, chickpea, lentil, faba bean, bread wheat, durum wheat.

This activity will strengthen crop improvement programs by suggesting efficient experimental designs to generate data on genotypes response, rationalization of test locations and identification of stable genotypes and locations. More detailed are given in the following.

**Objective: 1.** To develop experimental designs and statistical analysis of data to enhance the efficiency of crop germplasm development.

**Activities:** Use of efficient experimental designs and field plot techniques for varietal evaluation. Randomized plans for these designs will be generated. Statistical analysis of data from these designs will be carried out, including spatial adjustment. Computing codes will be written to obtain – estimate of error variance, test of significance on genotype means, efficiency of the design/method, BLUEs, BLUPs, genetic gain due to selection and realized gain, estimation of genetic parameters such as genotypic and phenotypic correlations, etc.

**Output:** Efficient designs, field plot techniques, and analyses used for the germplasm evaluation.

1. 100 variety trials analyzed for evaluation of blocking structure and spatial variability patterns.
2. Selected lines to move to the next year for evaluation and estimates of genetic gain

**Objective: 2.** To rationalize the evaluation environment for each crop lines, identify lines broad adaptation, identify lines for specific adaptation, identify stable locations, and quantify the rate of breeding progress and breeding success.

**Activities:** Compilation of data from each crop METs, individual trials BLUE and BLUP will be estimated, combined analysis of means will be performed to evaluate GxE interaction, various methods for stability analysis will be carried out, genotype x year interaction will be used to identify stable location, detection of any pattern in GxE interaction and/or G+GxE interaction will be examined and clustering of locations will be done on pattern predicted, the slope measuring the progress and success will be

estimated.

**Output:** Test locations for each crop identified, lines for broad adaptation, lines for specific adaptation, stable locations, and extent of breeding progress/success.

**Objective: 3.** To examine the effect of **climate change** on rationalization of the evaluation environment for each crop lines

**Activities:** Compilation of data from each crop METs and classified in contrasting climatic profiles or separated by time span. Within each of the two profiles, patterns in GxE interaction and/or G+GxE interaction will be examined and clustering of locations will be done on pattern predicted. Develop statistical tools to examine heterogeneity of the profiles for genotypic response to the environments.

**Output:** The change in the test locations for each crop identified for each climatic profile, lines for broad adaptation, lines for specific adaptation under each climatic profile.

**Objective: 4.** To share biometrical techniques and tools on exploitation of GxE interaction

**Activities:** Short term 2-3 weeks training courses will be conducted for NARSs scientists to share basic and advanced tools in biometrics and statistical computing.

**Output:** 100 scientists from NARS working in crop improvement trained in applications of Biometric tools

////////////////////////////////// EOF//////////////////////////////////