UTF/PAK/101/PAK

International FAO Training in Crop Yield Forecasting Modeling at University de Liege, Belgium

(31 July-September 8, 2007)

By

Muhammad Hayat Khan
Ijaz Ahmad
Muhammad Iftikhar Bhatti
Muhammad Asif
Ibrar-ul-Hassan Akhtar

Contacts

Imran Iqbal, SUPARCO
sgs@comsats.net.pk

René Gommes, FAO
rene.gommes@fao.org

Bernard Tychon, ULG
bernard.tychon@ulg.ac.be

Table of contents

1. <u>Introduction:</u>

The training was arranged by FAO for the first batch of fifteen trainees from 31st July 2007 to 8th September 2007 at University of Liege Arlon campus, Belgium. The said training is a component of foreign training of the approved PC-1 of the project, "Monitoring of Crops through Satellite Technology" under the UTF/PAK/101/PAK dated 20th September 2006 signed between Ministry of Food, Agriculture & Livestock and FAO of UN for transfer of technology. Objective of the training was to train officials from all stakeholders of the project in yield forecasting using FAO standard tools and methods.

All the participants arrived at the University's Arlon campus in a single group on 29th July 2007 from Islamabad via Istanbul, Turkey by same flight. Mr. Antoine, the training coordinator of University was present at the airport to receive and guide the group to Arlon, Campus.

<u>Objective</u>:

The training programme was based on general philosophy developed by FAO to estimate and forecast food production through integration of SRS and ground data at both data and product levels.

The programme was aimed to:

- Train participants in crop forecasting in general with available technical options and links between yield and area forecasting
- Familiarize them with different tools provided by FAO for crop monitoring and forecasting
- Adapt the general methodology to the Pakistani context

<u>Participants</u>:

1. Mr. Abdul Ghafoor	GM, SUPARCO
2. Mr. Muhammad Hayat Khan	Manager, SUPARCO
3. Mr. Khalid Latif	Manager, SUPARCO
4. Mr. Ijaz Ahmed Bhutta	Manger, SUPARCO
5. Mr. Muhammad Asif	Manager, SUPARCO
6. Mr. Iftikhar Ahmed Bhatti	Manager, SUPARCO
7. Miss Riffat Shamshad	AM, SUPARCO
8. Mr. Majid Masood Lodhi	AM, SUPARCO
9. Mr. Ibrar-ul-Hassan Akhtar	AM, SUPARCO
10. Mr. Muhammad Ashraf	CRS, Punjab
11. Mr. Noor-e-Islam	CRS, Sindh

12. Mr. Nasir Hameed	CRS, Balochistan, Quetta
13. Mr. Habib Ahmad	DD Economic Wing MinFAL
14. Mr.Waqar Aslam	Federal Bureau of Statistics
15. Mr.Qamar-ul-Zaman	Pak Met Deptt.

Presentation by Mr. Abdul Ghafoor GM SUPARCO:

A detail presentation was given by the group leader which included:

- Introduction to the project, its objectives and the progress made so far
- Introduction to agriculture in Pakistan regarding crop types, crop distribution, farming practices, irrigation system, weather conditions, main agriculture areas and the role of agriculture in Pakistan
- A brief on yield forecasting and area estimation techniques including general statistics and satellite based techniques
- Availability of input data required for yield forecast modeling in Pakistani context and included:
  - Geography of Pakistan
  - General population and %age of the population engaged in agriculture
  - Crop sequence
  - Crop management
  - Agriculture statistical data of the last 15 years from Min FAL

Presentation by Dr. Bernard Tychon (Training Supervisor):

Presentation by Dr. Bernard Tychon, Head Department of Environmental Sciences and Management University of Liege:

The presentation included:

- Introduction to Belgium including brief history, population, territorial components, agriculture etc. Belgium mainly has three regions, three communities and three languages (Dutch, French and German). Independent since 1830 with a population of 10.2 million.
- Detail presentation about the University of Liege as follows:
  - It was founded in 1817 and that it is the only public community sponsored University in French speaking part of the country.
  - It offers complete range of University courses at under graduate and post graduate levels with seven faculties; Philosophy and Letters, Law and School of Criminology,

Sciences, Medicines, Applied Sciences, Veterinary Medicines, Psychology and Educational Sciences. The University has a great repute and recognition at international level in the fields of Space Research and Aeronautics, Astronomy, Neurology, Microbiology and various branches of Biotechnology.

- The University comprises 3,300 employees with 2400 faculty members
- The Arlon campus comprises of faculty of sciences providing:
  - Post M.S and Ph. D in Environmental Sciences
  - Master Degree in Environmental Sciences and Management from 2007.
- The Arlon campus is known for its teaching and expertise in:
  - Agro-meteorology and processing of meteorological information
  - Natural risk management
  - Integrated management of water resources
  - Water technology
  - Monitoring of environmental variables
  - Sustainable socio economic development and environment
  - Public awareness of environment vulnerability
  - Waste management

2. Statistical aspects of training

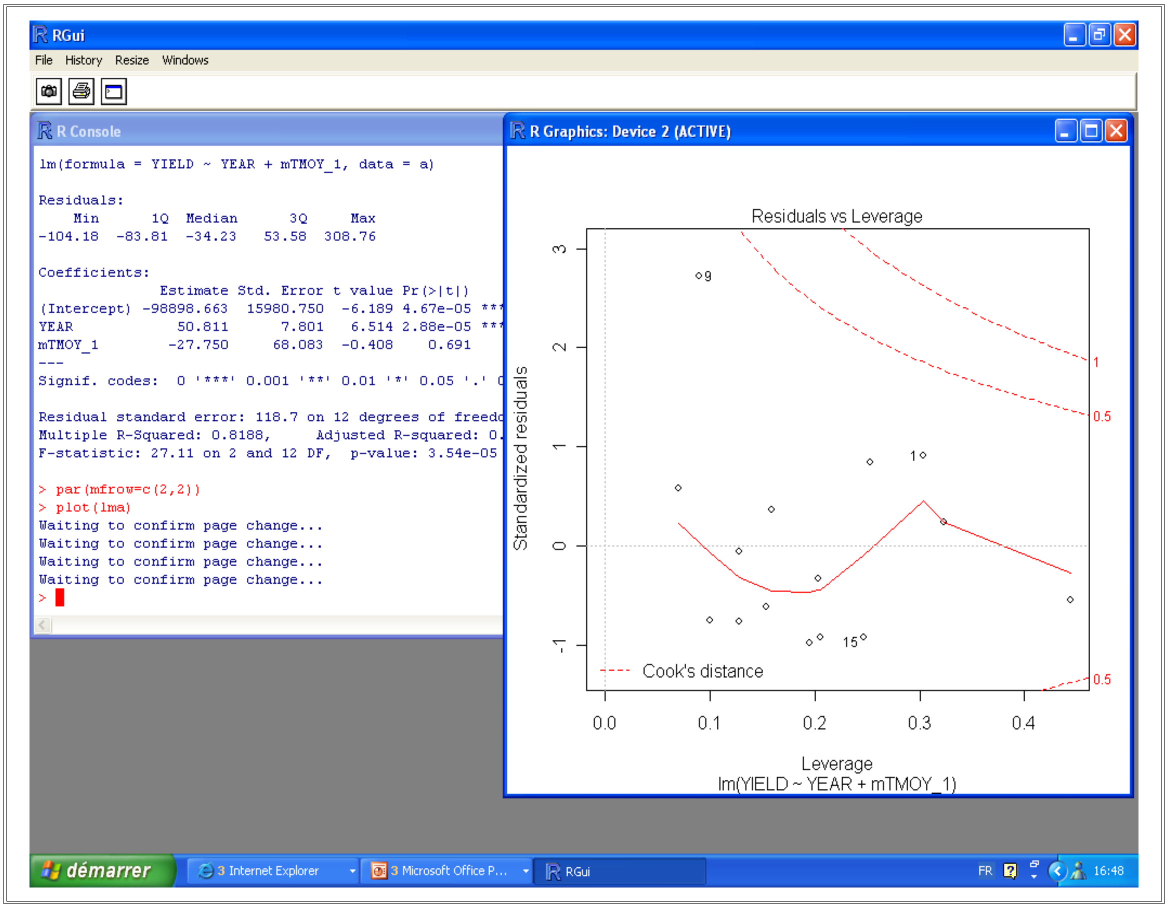
a. Data collection and arrangement

Data collection is the first and far most important step towards the development of cop yield forecasting model. Main aspect of data collection is the selection of the required input variables indispensable for crop yield models. This selection included mainly weather data i.e., maximum temperature, minimum temperature, rainfall, relative humidity, wind speed and sunshine hours or solar radiation, crop related data i.e., crop calendar, crop statistics for at least fifteen years, crop irrigation data, fertilizer data), satellite data i.e., complete series data of NDVI (Normalized Difference Vegetation Index) images. Moreover the daily data or at least 10 daily data is the minimum criteria to meet international and scientific standard for model development. Data collection sources in Pakistan could be Meteorological Department (PMD), MINFAL, and Irrigation Department etc.

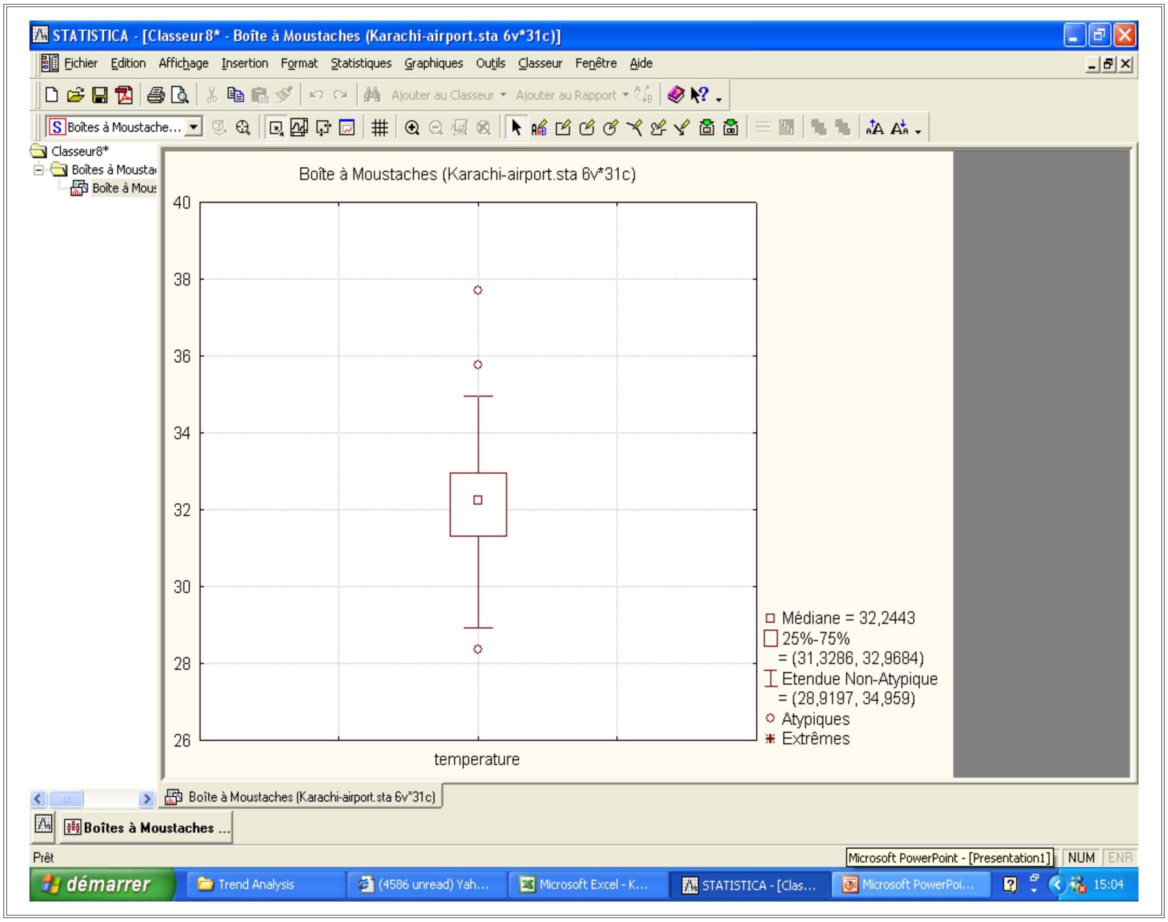
Data arrangement is the second most important step after data collection which mainly includes overview of collected data to view its standard of collection and to convert in into useable standard form. Best examples are cotton yield in bales and wind speed weather data which is mainly recorded in Knots where as for yield model we need to convert in into kg/ha and meter per second respectively.

b. Data quality control

Data quality control is check system for the evaluation of collected data before it's is used for model development. The only statistical way is the identification of "outliers" within collected data. This technique is mainly based on Cook distance and Whisker and Box plot tests.

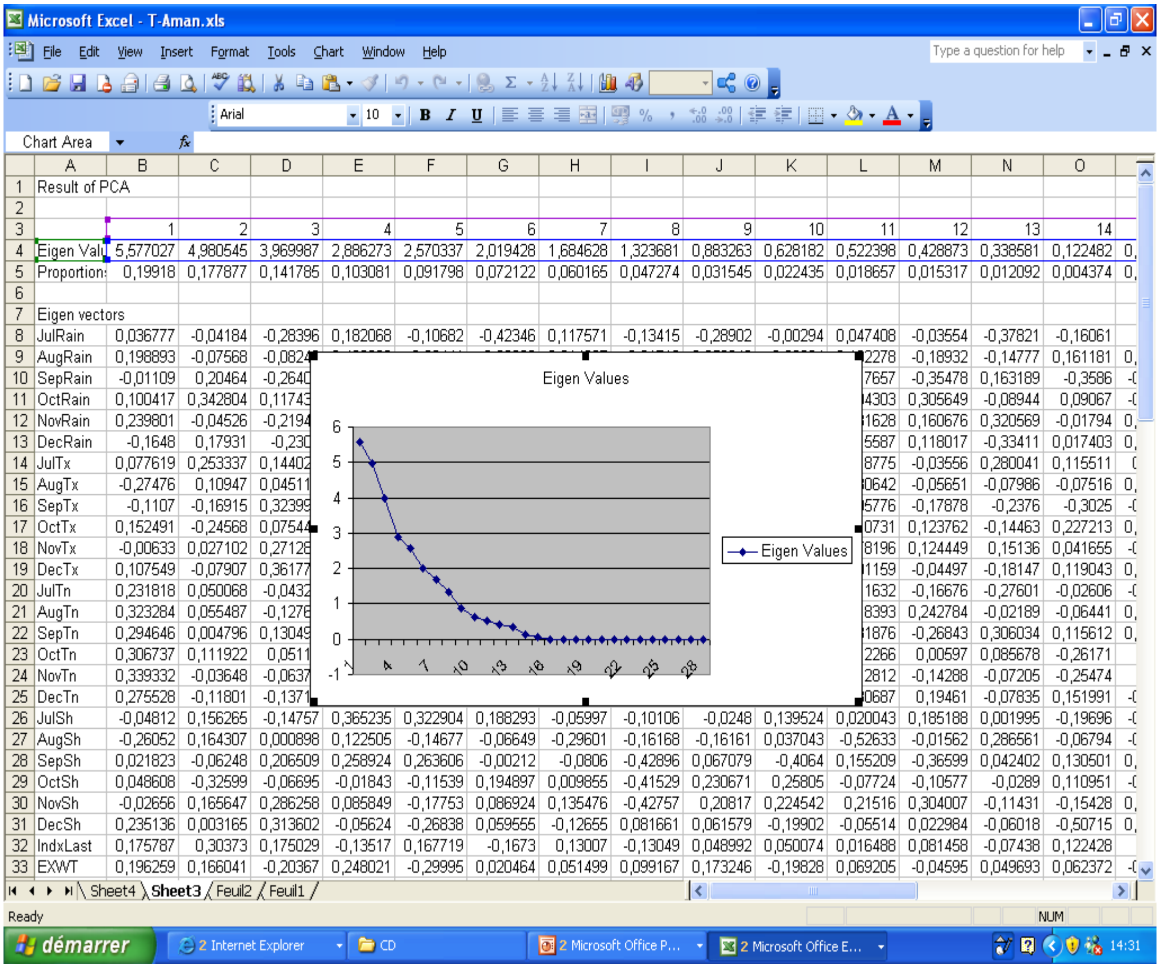

RGui
File History Resize Windows
R Console
lm(formula = YIELD ~ YEAR + mTMOY_1, data = a)
Residuals:
Min 1Q Median 3Q Max
-104.18 -83.81 -34.23 53.58 308.76
Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) -98898.663 15980.750 -6.189 4.67e-05 ***
YEAR 50.811 7.801 6.514 2.88e-05 ***
mTMOY_1 -27.750 68.083 -0.408 0.691

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.'
Residual standard error: 118.7 on 12 degrees of freedo
Multiple R-Squared: 0.8188, Adjusted R-squared: 0.
F-statistic: 27.11 on 2 and 12 DF, p-value: 3.54e-05
> par(mfrow=c(2,2))
> plot(lma)
Waiting to confirm page change...
Waiting to confirm page change...
Waiting to confirm page change...
Waiting to confirm page change...
R Graphics: Device 2 (ACTIVE)
Residuals vs Leverage
Standardized residuals
Cook's distance
Leverage
lm(YIELD ~ YEAR + mTMOY_1)
démarrer
3 Internet Explorer
3 Microsoft Office P...
RGui
FR
16:48

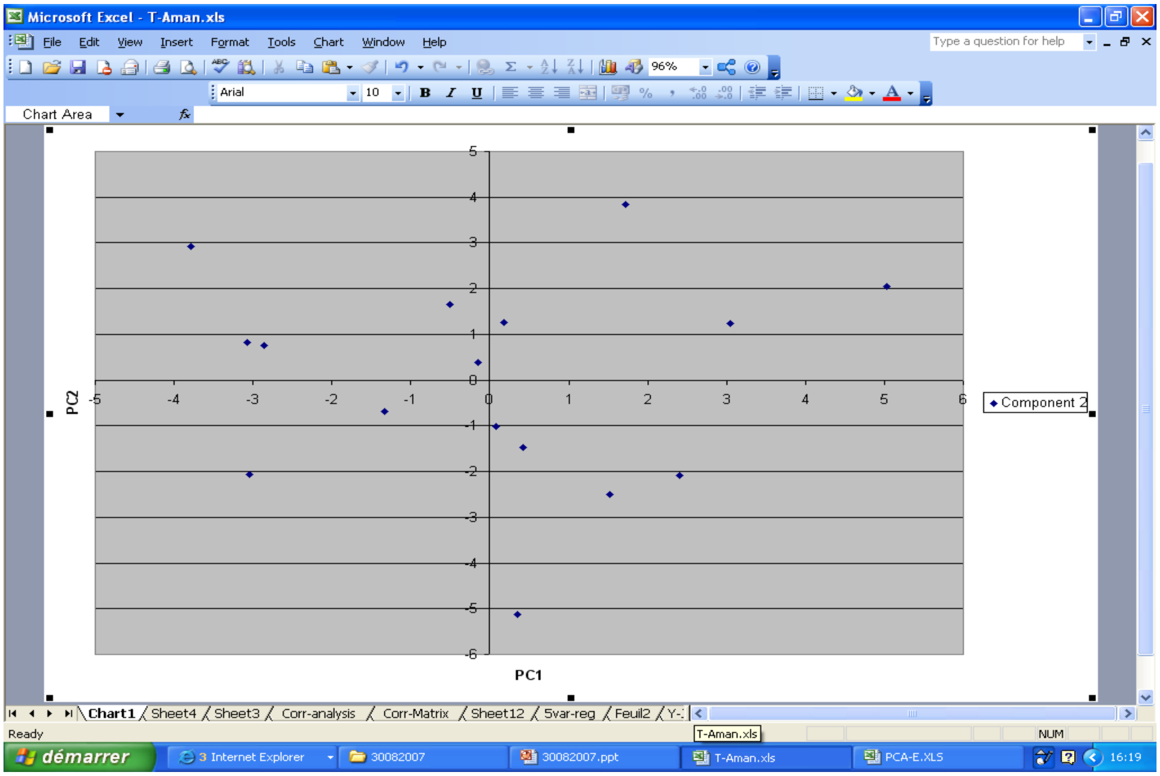


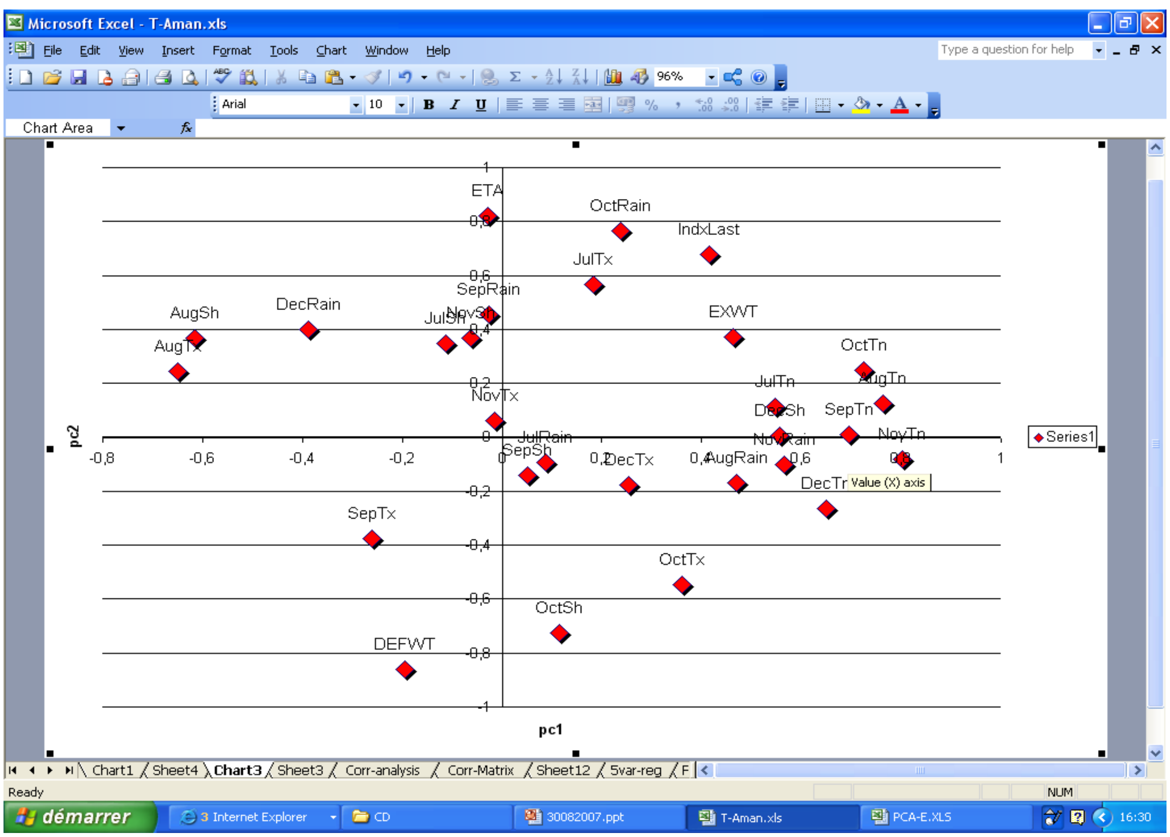
c. <u>**Principal Component Analysis**</u>

Principle component analysis is a way of identifying pattern in a given data and expressing the data in such a way as to highlight their similarities and differences. Since patterns in data is hard to find especially from high dimensional data set where graphical representation is not available. PCA is a powerful tool for analyzing data.



Other main advantage of PCA is that once you have found these patterns in the given data set and you compress the data by reducing the number of dimensions, without much loss of useful information. During cotton yield forecasting model calibration we produce 120 different explanatory variables like NDVI per decade, rainfall per decade, Evapo-Transpiration etc. We cannot easily identify the behavior of 120 different variables within and with other variables. We have two option first is the correlation matrix and second is the PCA. Correlation matrix gives just a single or one to one relation whereas PCA work in all dimensions.


Microsoft Excel - T-Aman.xls
PC2
PC1
Component 2
Chart1
Sheet4
Sheet3
Corr-analysis
Corr-Matrix
Sheet12
5var-reg
Feuil2
démarrer


Microsoft Excel - T-Aman.xls
ETA
OctRain
IndxLast
JulTx
SepRain
AugSh
DecRain
EXWT
AugTx
OctTn
JulTn
SepTn
NovTx
NovTn
JulRain
SepSh
DecTx
AugRain
DecTr
SepTx
OctTx
OctSh
DEFWT
pc2
pc1
Series1
Chart1
Sheet4
Chart3
Sheet3
Corr-analysis
Corr-Matrix
Sheet12
5var-reg
démarrer

d. Multiple regression analysis

The objective of multiple regression analysis is to predict the single dependent variable by a set of independent variables. There are some assumptions in using this statistics – (a) the criterion variable is assumed to be a random variable (b) there would be statistical relationship (estimating the average value) rather functional relationship (calculating an exact value) (c) there should be linear relationship among the predictors and between the predictors and criterion variable. Multiple regression analysis provides a predictive equation:

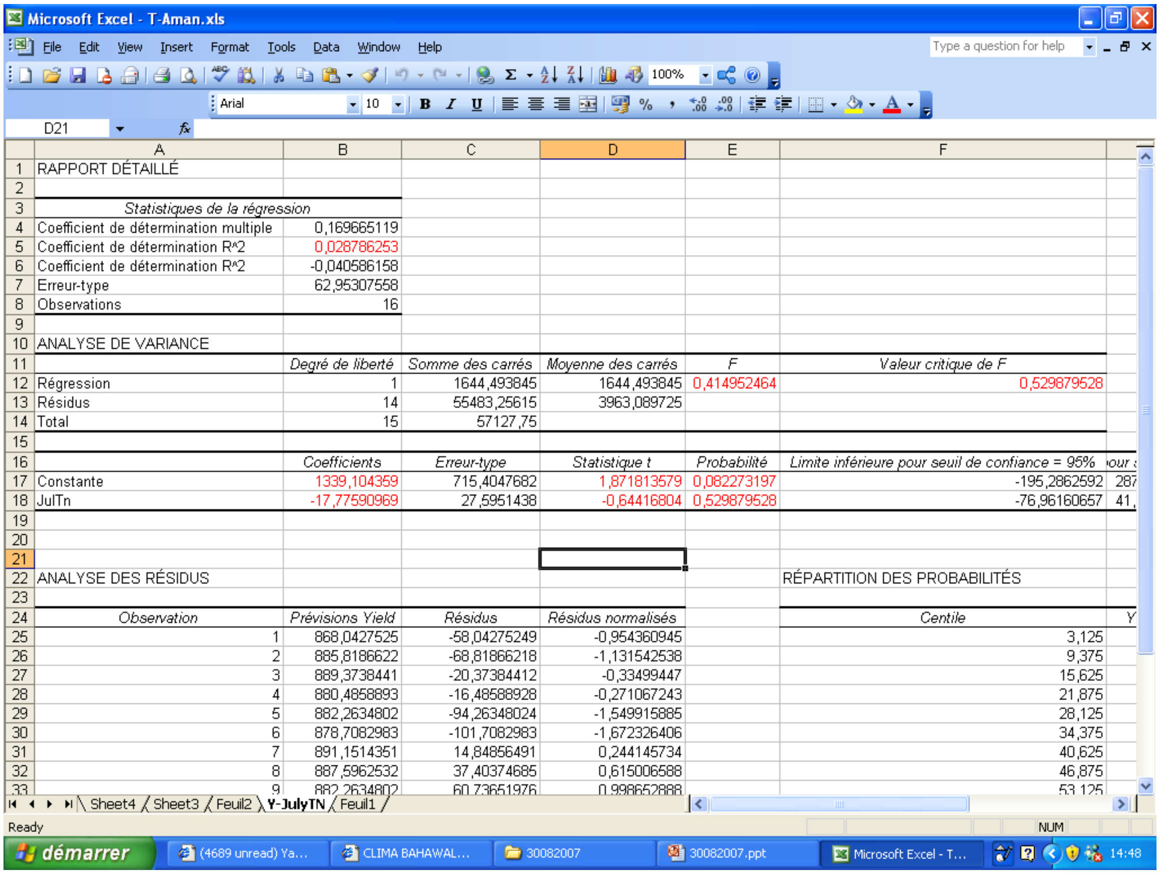
$Y = a + b1x1 + b2x2 + \ldots\ldots + bnxn$

Where, a = constant

b1, b2,… bn = beta coefficient or standardized partial regression coefficients (reflecting the relative impact on the criterion variable)

x1, x2,….. xn = scores on different predictors

The b's are the regression coefficients, representing the amount the dependent variable y changes when the corresponding independent changes 1 unit. The c is the constant, where the regression line intercepts the y axis, representing the amount the dependent y will be when all the independent variables are 0. The standardized version of the b coefficients is the beta weights, and the ratio of the beta coefficients is the ratio of the relative predictive power of the independent variables.



	A	B	C	D	E	F	
1	RAPPORT DÉTAILLÉ						
2							
3	*Statistiques de la régression*						
4	Coefficient de détermination multiple	0,169665119					
5	Coefficient de détermination R^2	0,028786253					
6	Coefficient de détermination R^2	-0,040586158					
7	Erreur-type	62,95307558					
8	Observations	16					
9							
10	ANALYSE DE VARIANCE						
11		*Degré de liberté*	*Somme des carrés*	*Moyenne des carrés*	*F*	*Valeur critique de F*	
12	Régression	1	1644,493845	1644,493845	0,414952464	0,529879528	
13	Résidus	14	55483,25615	3963,089725			
14	Total	15	57127,75				
15							
16		*Coefficients*	*Erreur-type*	*Statistique t*	*Probabilité*	*Limite inférieure pour seuil de confiance = 95%*	*our*
17	Constante	1339,104359	715,4047682	1,871813579	0,082273197	-195,2862592	287
18	JulTn	-17,77590969	27,5951438	-0,64416804	0,529879528	-76,96160657	41,
19							
20							
21							
22	ANALYSE DES RÉSIDUS					RÉPARTITION DES PROBABILITÉS	
23							
24	*Observation*	*Prévisions Yield*	*Résidus*	*Résidus normalisés*		*Centile*	*Y*
25	1	868,0427525	-58,04275249	-0,954360945		3,125	
26	2	885,8186622	-68,81866218	-1,131542538		9,375	
27	3	889,3738441	-20,37384412	-0,33499447		15,625	
28	4	880,4858893	-16,48588928	-0,271067243		21,875	
29	5	882,2634802	-94,26348024	-1,549915885		28,125	
30	6	878,7082983	-101,7082983	-1,672326406		34,375	
31	7	891,1514351	14,84856491	0,244145734		40,625	
32	8	887,5962532	37,40374685	0,615006588		46,875	
33	9	882,2634802	60,73651976	0,998652888		53,125	

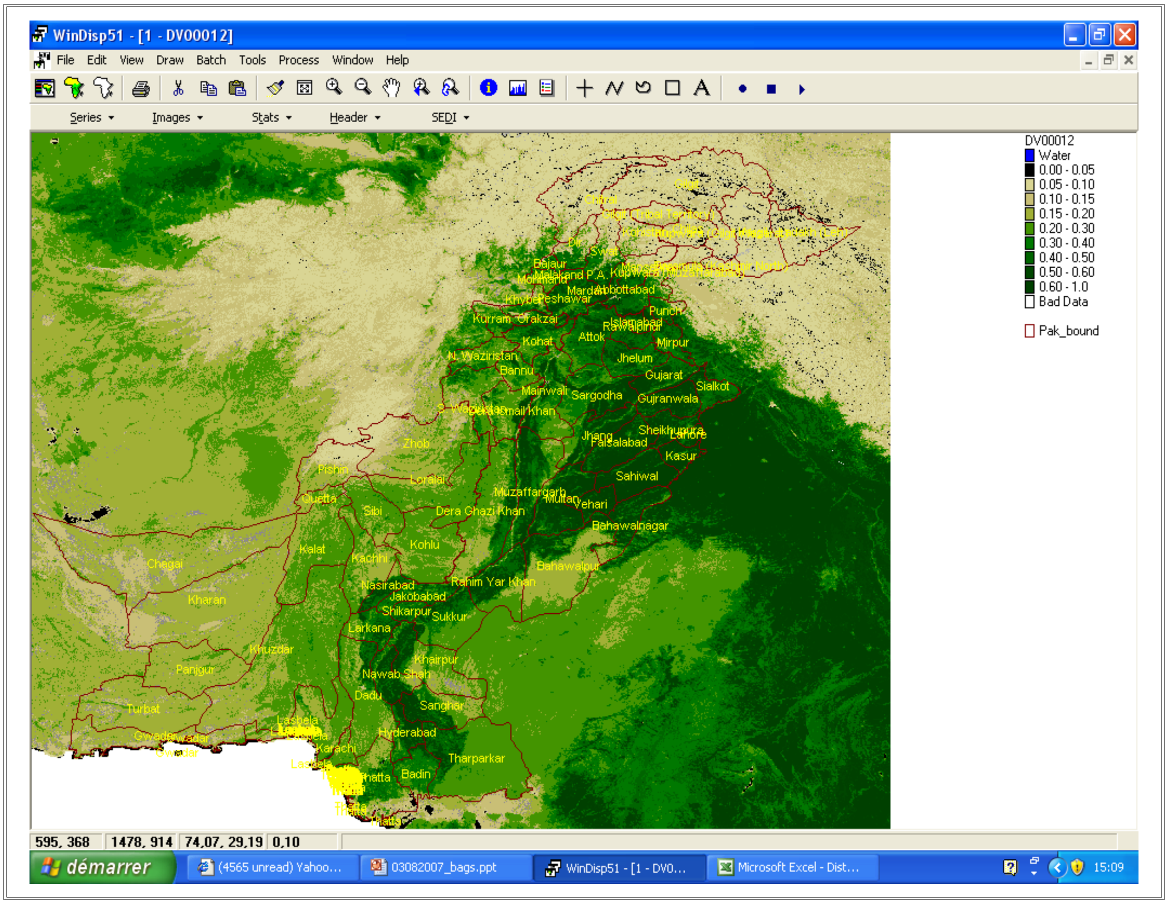
Associated with multiple regressions is R2, multiple correlations, which is the percent of variance in the dependent variable, explained collectively by all of the independent variables. Multiple regression shares all the assumptions of correlation: linearity of relationships, the same level of relationship throughout the range of the independent variable ("homoscedasticity"), interval or near-interval data, absence of outliers, and data whose range is not truncated. In addition, it is important that the model being tested is correctly specified. The exclusion of important causal variables or the inclusion of extraneous variables can change markedly the beta weights and hence the interpretation of the importance of the independent variables.

3. <u>Software Pool in relation to Crop Yield Forecasting Model</u>

There was a very comprehensive list of software's that we have used or learn about their utility during training at ULG. This list mainly include WinDisp, Vast, Bags, Land Cover Classification System (LCCS), Geo-Vis, ADG, Budget, Rainbow, AMS, R-Console and different conversion software's. But main core software's necessary for Crop yield Model developments are;

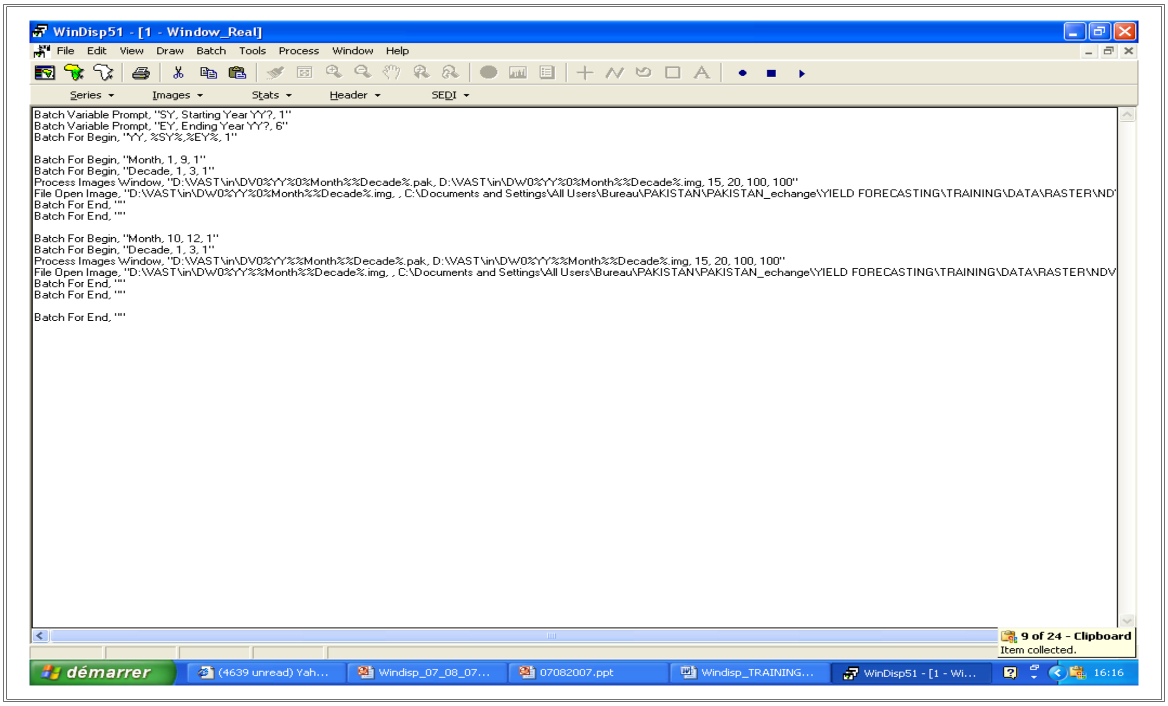
a. **<u>WinDisp:</u>**

Windisp is Image display and analysis software. It was mainly used for NDVI images analysis during training. We can handle a set of data in two ways mainly



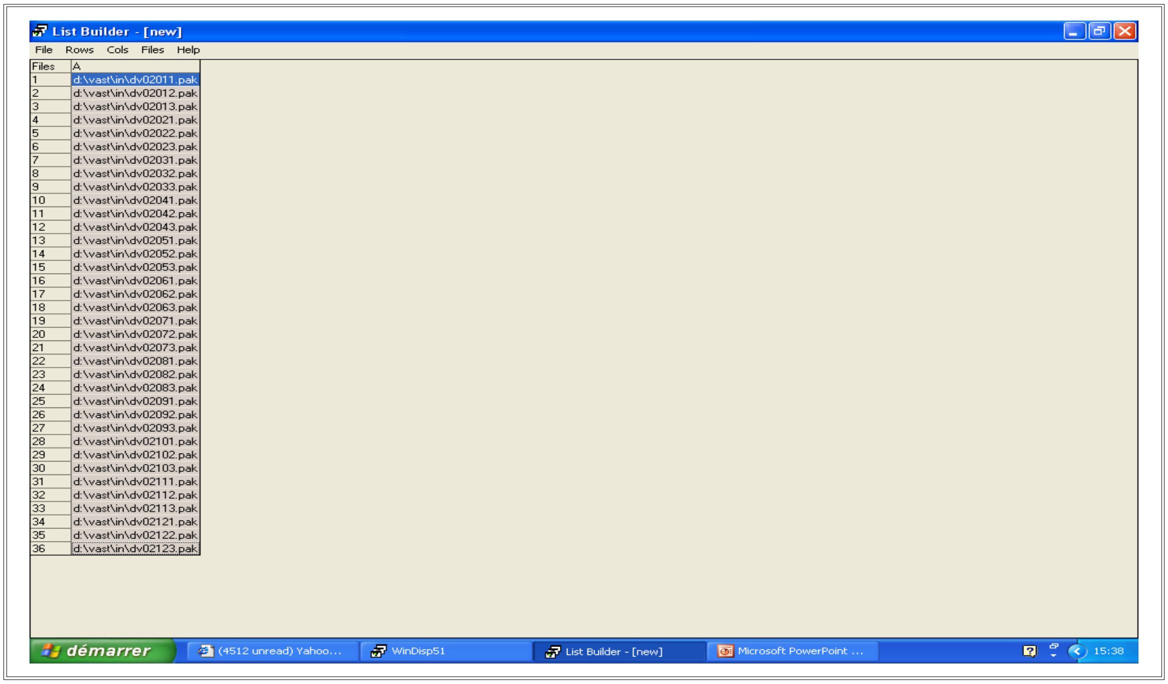
i. <u>Batch processing:</u>

In batch processing, we make a recording of what method we used for the required output once and then we run it to all through editing of batch file recorded earlier which make the process time saving with standard methodology.

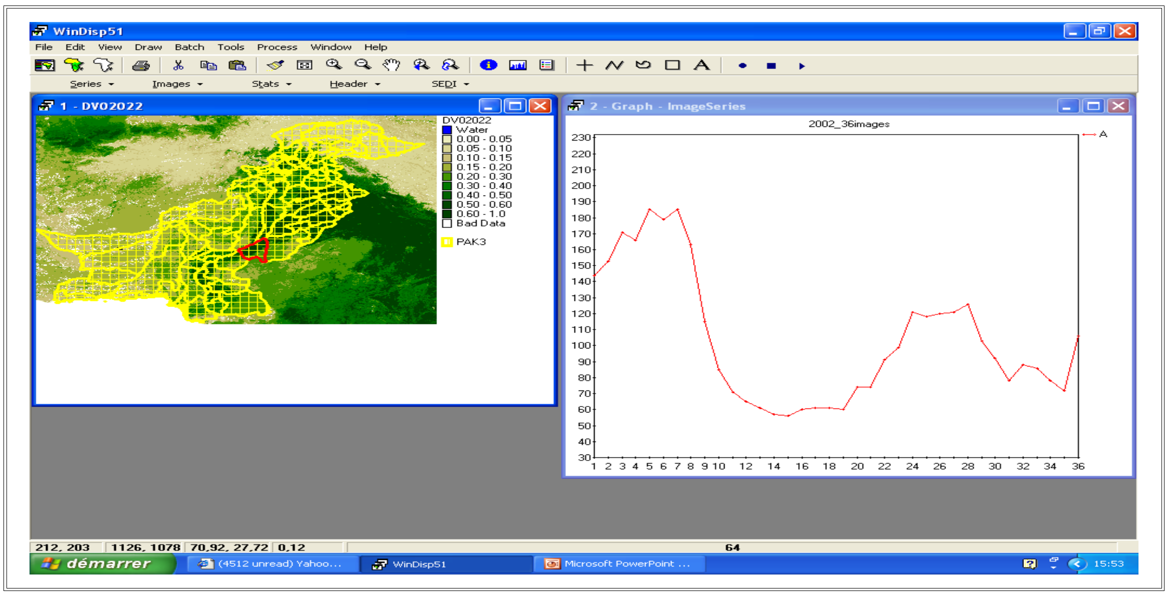


ii. <u>Image list method:</u>

In image list, we enter all the series of images to be processed in a newly created file list in WinDisp and then saved. This is very useful when we have to process NDVI images for a particular crop like wheat starting from November and end in May.



The important aspect of software is the "process tool" through which we worked on series of NDVI images and statistical extraction from NDVI images. In series option we get an output as image whereas in other case we get only statistical information available in input images. We get mean, median, minimum, maximum, sum, count, slope, standard deviation, minimum date and maximum date output in the form of an image developed from series of images. Whereas in case of statistical processing we get median, minimum, maximum, range, count, standard deviation from an image (NDVI image).



This is also important to process the image output from other software's i.e., VAST and BAGS.

b. VAST:

Vegetation Analysis in Space and Time (VAST) is a simple program written in 1994 which extracts some useful information from temporal satellite NDVI image series at pixel level. This software is command based which extract directly from images in a defined folder. This software automatically processed command from input folder into output folder.

Command of the VAST is mainly consist of two essential and seven optional parameters.

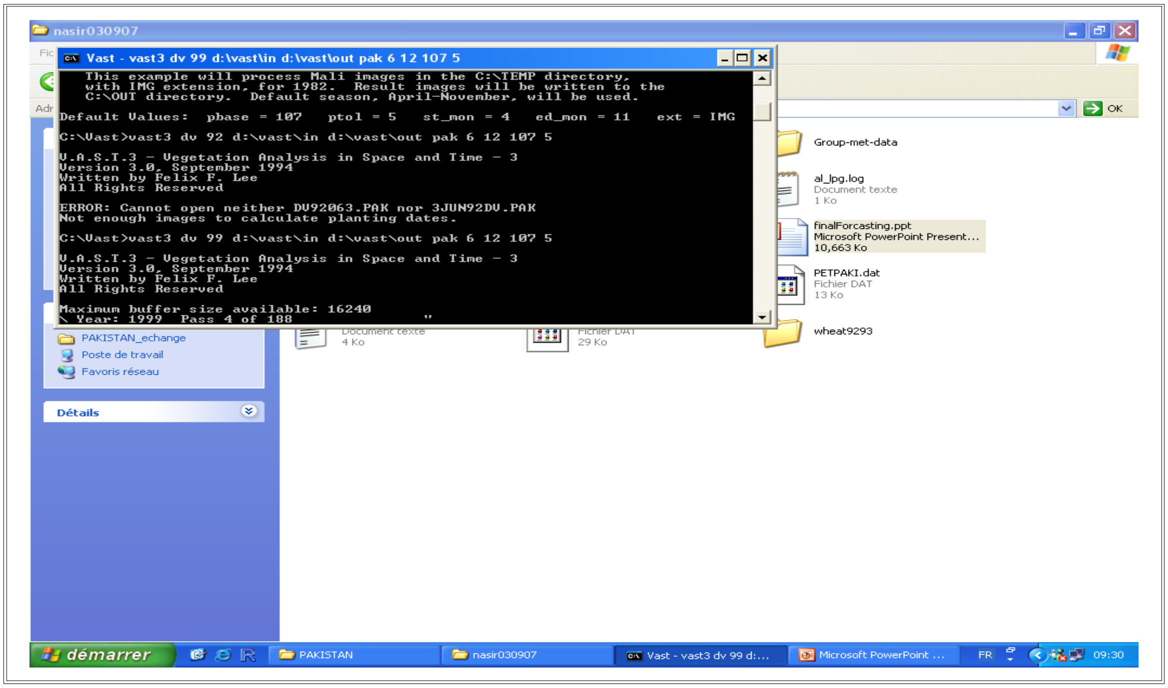
Essential parameters:

(1) Country code

(2) Year

Optional parameters:

(1) Input directory
(2) Output directory
(3) File extension
(4) Crop starting decade month
(5) Crop ending decade month
(6)Base NDVI value for start of crop
(7)Minimum NDVI increase for season start for a given crop



Command is VAST3 PK 99 c\vast\in c\vast\out img 5 11 130 05

Execution of VAST command give 12 different output from series of NDVI images for a given crop season like June to November for cotton and important ones are as follows;

a. Crop starting Decade (SDAT)
b. Crop starting NDVI value for starting decade (SVAL)
c. Crop ending NDVI value for ending decade (EVAL)
d. Crop peak growth decade (PEAK)
e. Crop peak growth NDVI value (PVAL)
f. Cumulative NDVI for whole crop season (CUMM)
g. DROP = Peak NDVI value-Ending NDVI value
h. SLOPE = Linear growth of crop from germination to peak growth stage (SDAT,SVAL) to (PEAK, PVAL)
i. SKEW = sum of three NDVI after PEAK (PEAK1+2+3)/sum of the seven NDVI values (Peak-3-2-1+peak+Peak1+2+3)
j. HORZ = PEAK decade – Starting decade (SDAT)

k. VERT = Peak NDVI value – Starting NDVI value

These outputs were in image format which required WinDisp software for the extraction of statistical information in relation to above mentioned output images. These outputs make a very strong basis about crop development behaviour in the field through NDVI images.

c. <u>BAGS</u>

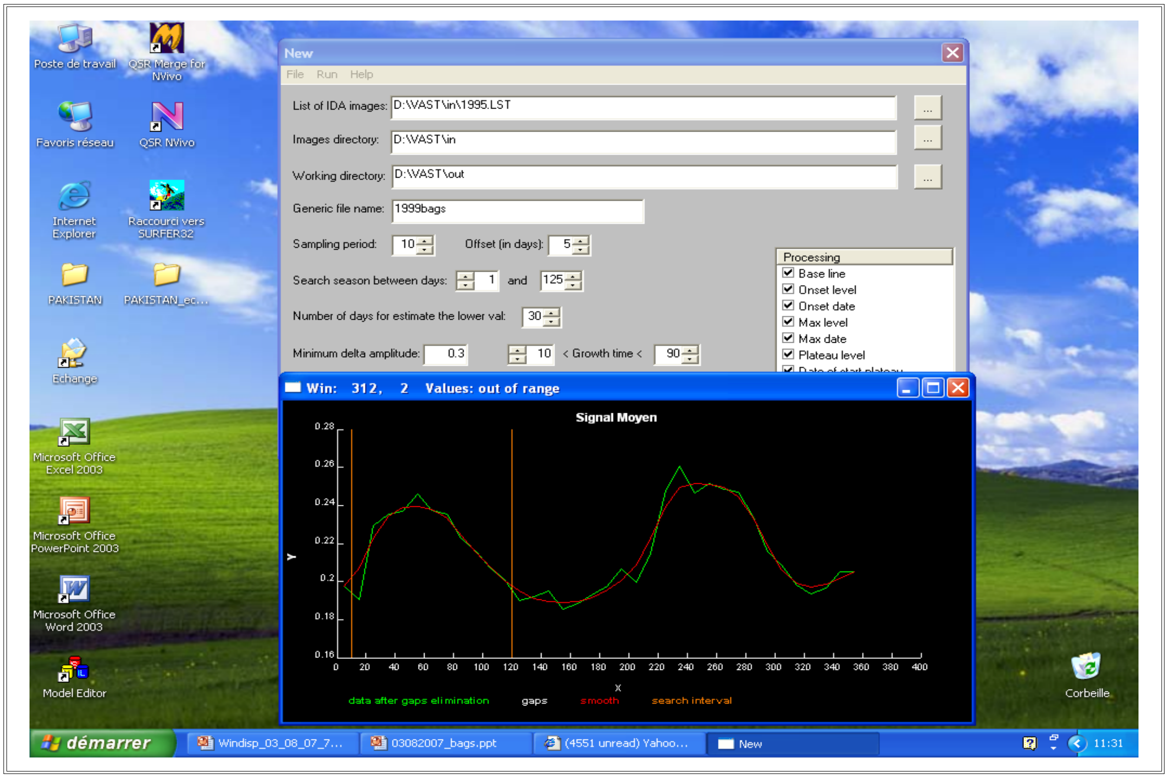
BAGS software work similar to VAST in which it gives different outputs related to a given crop based on the NDVI images. This required some very important inputs for image analysis which includes:

(1) Sampling period like 10 daily
(2) Season search between particular days of the year like 120th day of year to 300th day of the year
(3) Number of days to estimate lowest value for a season
(4) Average crop growth period in days
(5) Starting and ending NDVI value

BAGS give eight different outputs from series of NDVI images for a given crop season like June to November for cotton and important ones are as follows:

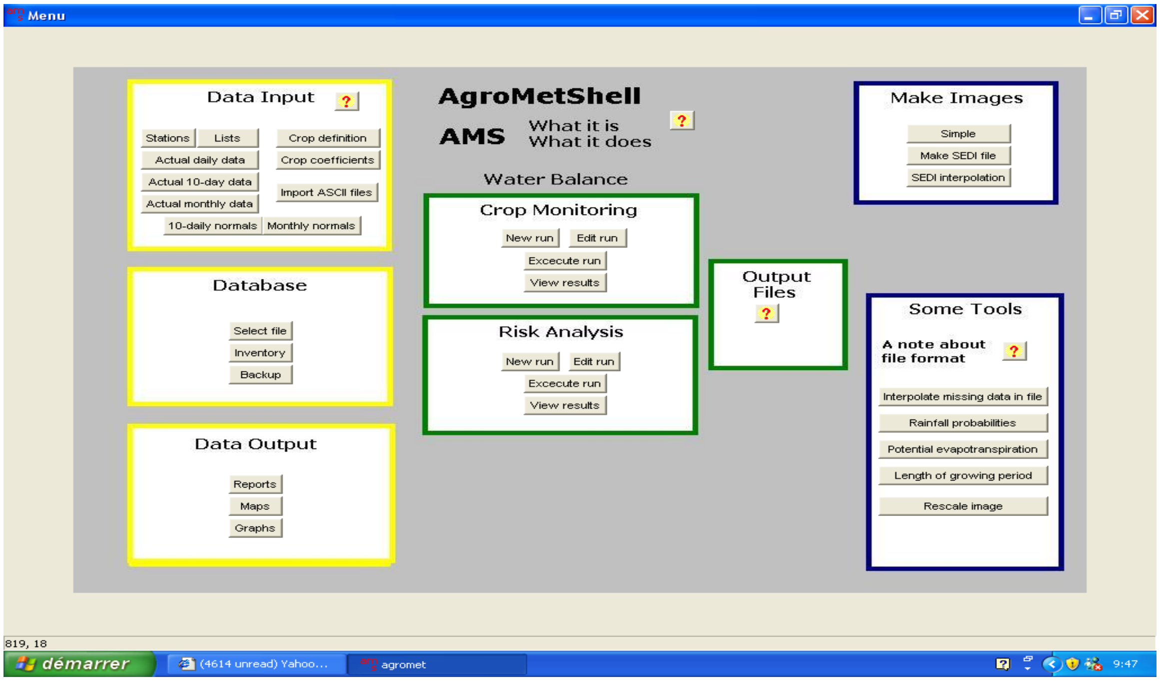
(1) Crop onset date
(2) Crop onset Level
(3) Crop NDVI maximum Date
(4) Crop NDVI maximum Level
(5) Baseline
(6) Plateau Level
(7) Date of start of plateau level
(8) Date of end of plateau level

These outputs were in image format which required WinDisp software for the extraction of statistical information in relation to above mentioned output images. These outputs make a very strong basis about crop development behaviour in the field through NDVI images.



d. Agro-Met Shell (AMS)

AgroMetShell provides a toolbox for agro-meteorological crop monitoring and forecasting. This software includes a database that holds all weather, climate and crop data needed to analyze weather impacts on crops. AMS follows the FAO philosophy of crop specific soil water balance (CSSWB) which mainly stress the analysis of possible impact weather data on crop in combination of crop characteristics, water, fertilizer input etc.



e. <u>R-Console</u>

R is a language and environment for statistical computing and graphics. R provides a wide variety of statistical (linear and nonlinear modeling, 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.

One of R's strengths is the ease with which well-designed publication-quality plots can be produced, including mathematical symbols and formulae where needed. Great care has been taken over the defaults for the minor design choices in graphics, but the user retains full control

R is an integrated suite of software facilities for data manipulation, calculation and graphical display. It includes:

- An effective data handling and storage facility,
- A suite of operators for calculations on arrays, in particular matrices,
- A large, coherent, integrated collection of intermediate tools for data analysis,
- Graphical facilities for data analysis and display either on-screen or on hardcopy, and

- A well-developed, simple and effective programming language which includes conditionals, loops, user-defined recursive functions and input and output facilities.

The term "environment" is intended to characterize it as a fully planned and coherent system, rather than an incremental accretion of very specific and inflexible tools, as is frequently the case with other data analysis software.

4. Technical Analyses for Model development

α. AgroMetShell Analysis

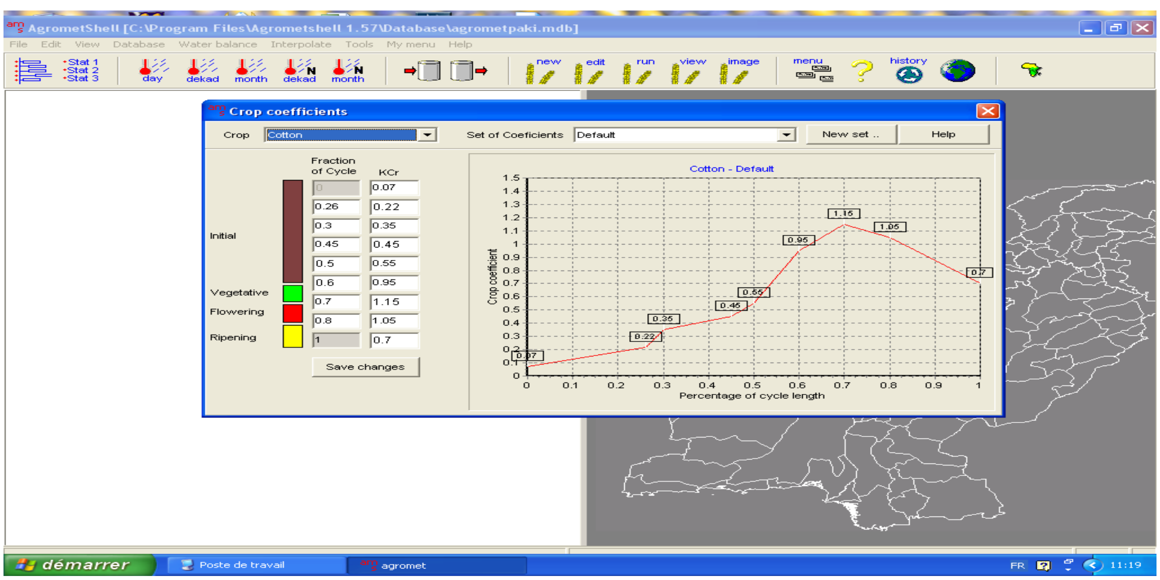
The important aspects for AMS are as follows:

(1) It needs the data in comma separated data file “CSV” format for the database development.

(2) Data which it can handle and process mainly include actual daily, 10 daily and monthly data from where we can generate normalized data.

(3) We have to develop proper crop coefficients to get proper crop specific outputs like Evapo-transpiration, water deficit etc

(4) We have to define proper parameters for crop duration and irrigation parameters for final agro-meteorological outputs.



Main explanatory variable extracted through AMS;

i. Crop specific Harvest indices (Normal, Latest)
ii. Water in Excess at specific crop development stage
iii. Water in Deficit at specific crop development stage
iv. Total crop water requirement
v. Evapo-Transpiration at specific crop stage

β. **Statistical Analysis:**

All the available data like meteo-data (maximum temperature, minimum temperature, relative humidity, rainfall, wind speed and solar radiation i.e. 10 daily), NDVI images, AMS outputs, VAST extracted variables and WinDisp statistical NDVI extraction variables were arranged in excel under the umbrella of "Crop Yield Model Explanatory Variables". Main variables used were as follows

Decadal rainfall	Crop water satisfaction index	Peak NDVI
Decadal accumulated Rainfall	Crop water deficit at different crop stages	Peak NDVI Date
Decadal NDVI	Crop water excess at different crop stages	Other---------------
Decadal accumulated NDVI	Crop specific Evapo-Transpiration	
Crop water requirement	Starting Decade	
Crop Harvest index	Starting Decade NDVI	

In R-software we open the explanatory variable and applied following statistical test

(1) Correlation matrix
(2) Multiple regression analysis

For this analysis we use three functions in R-console on all data sets which are:

(1) lm
(2) mle.stepwise
(3) stepAIC

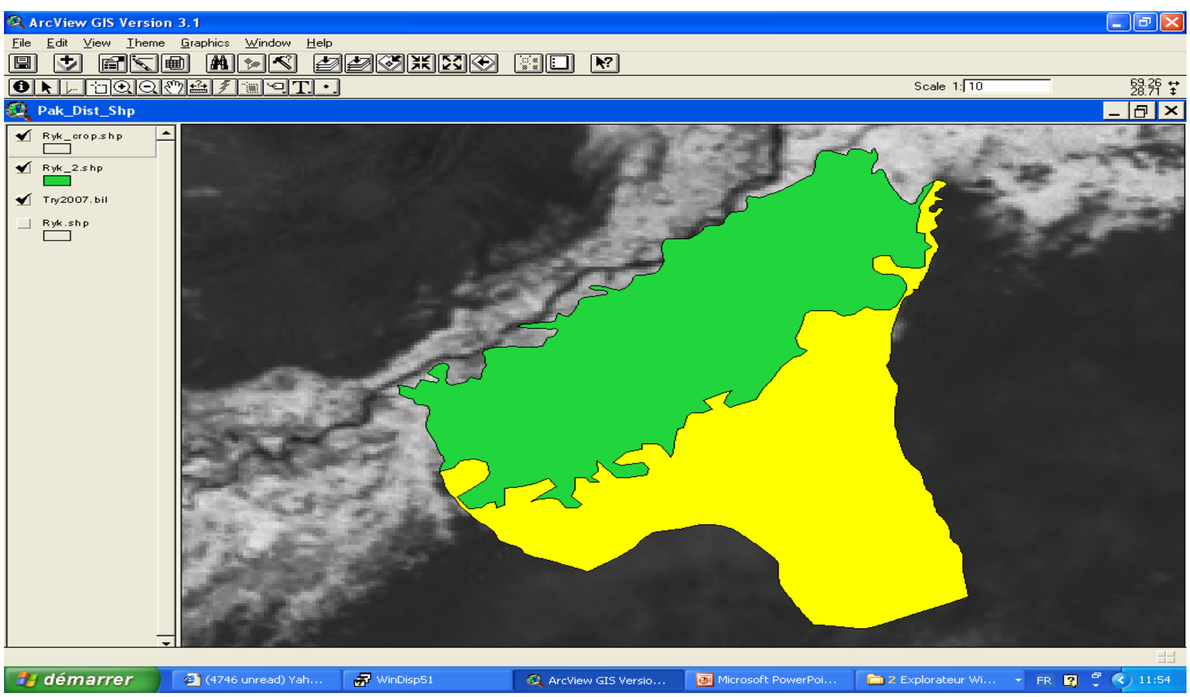
These three functions results a calibrated yield forecast model with possible significant predictor variables.

5. Crop Yield Forecasting Models

Three crop yield forecasting model were developed by three groups at University of Leige, Arlon, Belgium. These models were

a. Cotton Model for Rahim Yar Khan
b. Cotton Model for Faisalabad
c. Wheat Model for Sialkot

Districts	Crop	Yield Models
Faisalabad	Cotton	292,138 + (Cummulative Rainfall from last decade of May to second Decade of August*1,647 - (Cummulative Rainfall from last decade of May to last Decade of August*1,009
Rahim Yar Khan	Cotton	812.068 – (ETA at Ripening * 2.981)
Sialkot	Wheat	78252.50 - (YEAR*40.389)+ (-581.0304*April 3rd decade NDVI)



ile Edit View Misc Packages Windows Help

```
yk<-read.table("r-file.csv",header=T)
or in file(file, "r") : unable to open connection
addition: Warning message:
not open file 'r-file.csv', reason 'No such file or directory' in: file(file, "r")
yk<-read.table("r-file.csv",header=T)
or in file(file, "r") : unable to open connection
addition: Warning message:
not open file 'r-file.csv', reason 'No such file or directory' in: file(file, "r")
yk<-read.table("r-file.csv",header=T)
yk
or in do.call("data.frame", rval) : symbol print-name too long
yk<-read.table("r-file.txt",header=T)
yk
Year NDVI.16 NDVI.17 NDVI.18 NDVI.19 NDVI.20 NDVI.21 NDVI.22 NDVI.23 NDVI.24 NDVI.25 NDVI.26 NDVI.27 NDVI.28 NDVI.29
1992   0.180   0.190   0.200   0.180   0.200   0.220   0.330   0.395   0.460   0.530   0.580   0.630   0.620   0.590
1993   0.210   0.190   0.170   0.230   0.295   0.360   0.400   0.430   0.460   0.520   0.560   0.600   0.620   0.575
1994   0.190   0.185   0.180   0.190   0.235   0.280   0.410   0.480   0.550   0.580   0.615   0.650   0.620   0.570
1995   0.150   0.140   0.130   0.180   0.215   0.250   0.300   0.340   0.380   0.550   0.590   0.630   0.590   0.545
1996   0.160   0.155   0.150   0.180   0.220   0.260   0.370   0.420   0.470   0.490   0.510   0.530   0.510   0.505
1997   0.170   0.165   0.160   0.220   0.245   0.270   0.350   0.395   0.440   0.500   0.530   0.560   0.560   0.555
1998   0.190   0.215   0.240   0.270   0.290   0.310   0.350   0.390   0.430   0.510   0.545   0.580   0.590   0.560
1999   0.152   0.188   0.196   0.272   0.332   0.344   0.480   0.508   0.536   0.524   0.536   0.524   0.460   0.448
2000   0.132   0.148   0.164   0.224   0.252   0.364   0.388   0.448   0.496   0.476   0.504   0.476   0.468   0.460
2001   0.120   0.132   0.156   0.252   0.296   0.420   0.440   0.584   0.532   0.516   0.536   0.544   0.492   0.444
2002   0.156   0.188   0.188   0.212   0.312   0.340   0.404   0.412   0.500   0.456   0.484   0.480   0.480   0.420
2003   0.148   0.176   0.216   0.404   0.468   0.504   0.516   0.524   0.648   0.580   0.600   0.580   0.524   0.460
2004   0.148   0.164   0.252   0.268   0.328   0.388   0.448   0.448   0.476   0.532   0.500   0.484   0.528   0.404
2005   0.184   0.208   0.240   0.364   0.380   0.416   0.464   0.500   0.536   0.564   0.592   0.556   0.568   0.496
NDVI.30 NDVI.31 NDVI.32 NDVI.33 NDVI.34 Cum.NDVI1 Cum.NDVI2 Cum.NDVI3 Cum.NDVI4 Cum.NDVI5 Cum.NDVI6 Cum.NDVI7 Cum.NDVI8
  0.560   0.490   0.455   0.420   0.310     0.370     0.570     0.750     0.380     0.600     0.930     0.725     1.185
  0.530   0.440   0.405   0.370   0.280     0.400     0.570     0.800     0.525     0.885     1.285     0.830     1.290
  0.520   0.420   0.395   0.370   0.350     0.375     0.555     0.745     0.425     0.705     1.115     0.890     1.440
  0.500   0.440   0.410   0.380   0.340     0.290     0.420     0.600     0.395     0.645     0.945     0.640     1.020
  0.500   0.490   0.465   0.440   0.410     0.315     0.465     0.645     0.400     0.660     1.030     0.790     1.260
  0.550   0.520   0.495   0.470   0.360     0.335     0.495     0.715     0.465     0.735     1.085     0.745     1.185
  0.530   0.470   0.450   0.430   0.380     0.405     0.645     0.915     0.560     0.870     1.220     0.740     1.170
  0.400   0.368   0.340   0.324   0.264     0.340     0.536     0.808     0.604     0.948     1.428     0.988     1.524
  0.416   0.376   0.352   0.308   0.308     0.280     0.444     0.668     0.476     0.840     1.228     0.836     1.332
  0.416   0.364   0.340   0.312   0.264     0.252     0.408     0.660     0.548     0.968     1.408     1.024     1.556
```

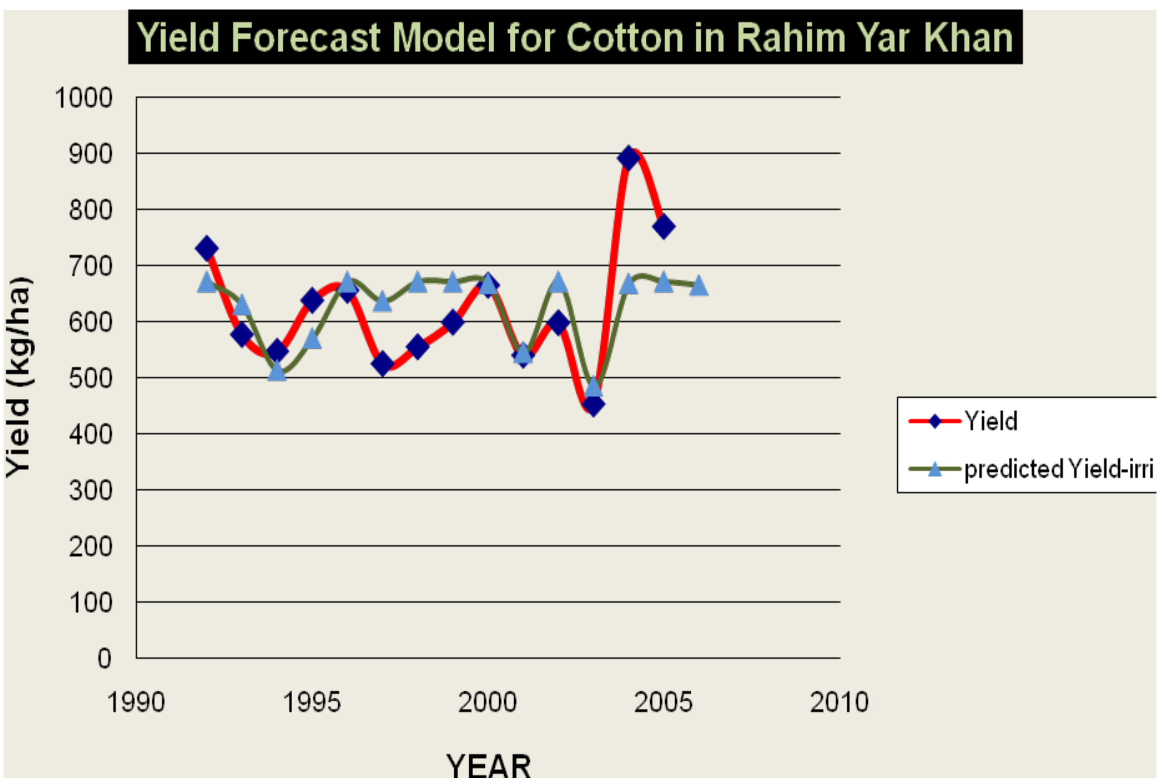
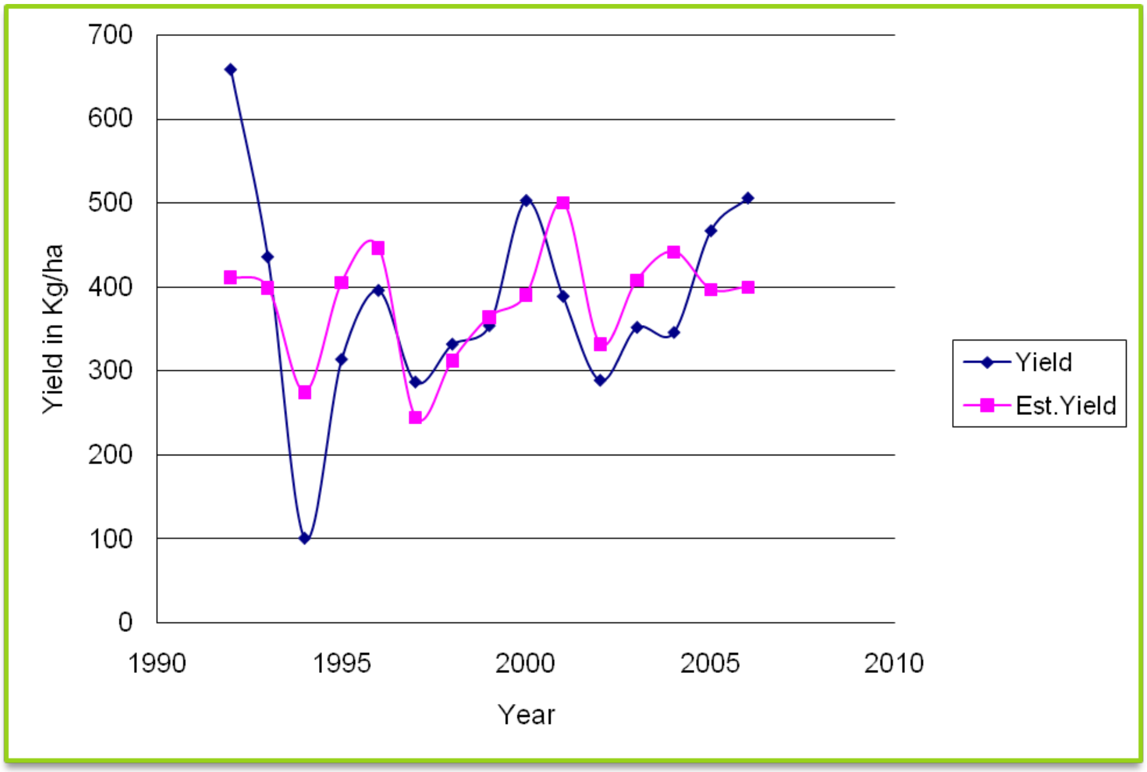
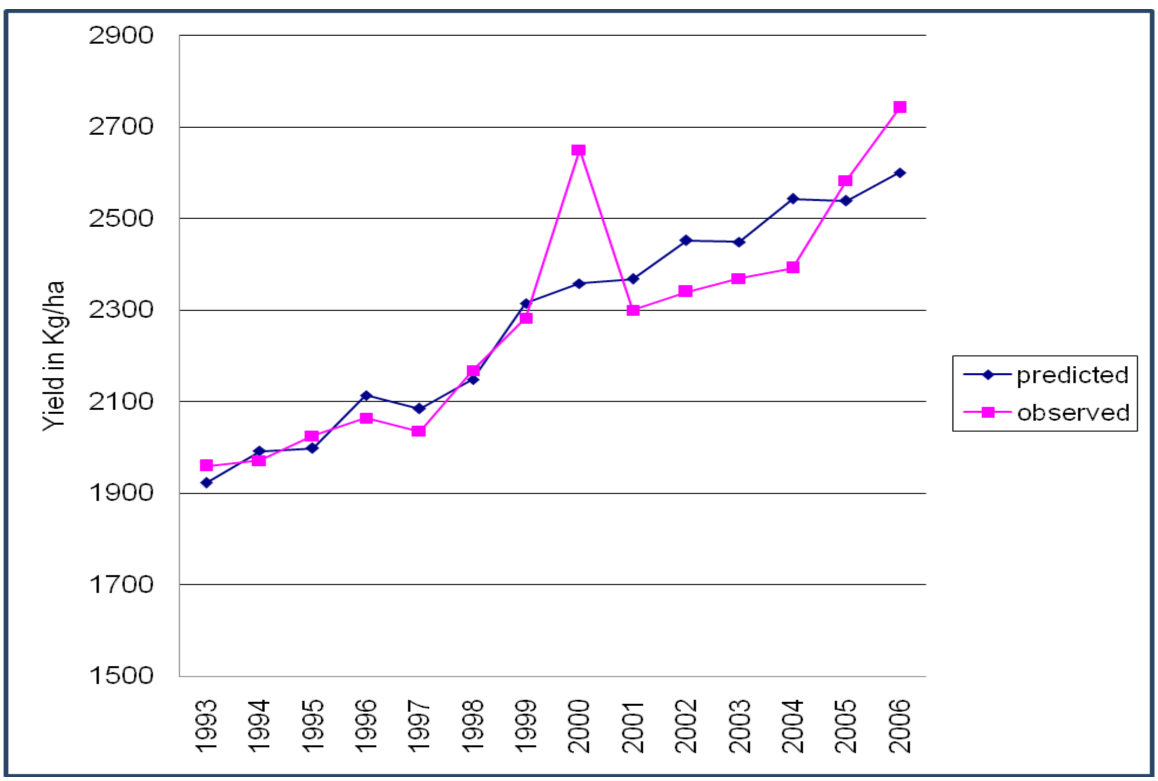


Table: Observed vs model predicted yield Forecast from Yield

YEAR	RYK Observed YIELD (cotton)	RYK PREDICTED YIELD (cotton)	Faisalabad Observed YIELD (cotton)	Faisalabad PREDICTED YIELD (cotton)	Sialkot Observed YIELD (Wheat)	Sialkot PREDICTED YIELD (Wheat)
1992	730	670.298	659	411.5716	----	----
1993	576	629.335	436	399.2582	1960	1923
1994	546	512.748	101	275.0484	1972	1993
1995	637	569.466	314	405.6398	2026	1998
1996	655	670.298	396	446.8125	2065	2114
1997	524	635.637	287	245.2006	2036	2085
1998	554	670.298	332	312.8954	2169	2149
1999	598	670.298	354	364.551	2283	2316
2000	664	667.147	503	390.6865	2651	2358
2001	539	544.258	389	500.126	2300	2368
2002	597	670.298	289	332.1994	2340	2453
2003	452	484.389	352	408.2202	2370	2449
2004	891	667.147	346	442.0466	2392	2543
2005	769	670.298	467	397.3394	2584	2539
2006	766	663.996	506	400.0982	2742	2601


700
600
500
400
300
200
100
0
1990
1995
2000
2005
2010
Year
Yield in Kg/ha
Yield
Est.Yield


2900
2700
2500
2300
2100
1900
1700
1500
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
Yield in Kg/ha
predicted
observed

6. <u>Observations:</u>

(1) So far the overall training is concerned; it was good, conceptual with regard to yield forecast modeling and practical in nature.

(2) All the requisite statistical data for the last fifteen years with all necessary variables be obtained from the concerned departments and compiled on the required format for use in modeling to the Pakistani context. The data include:

a) **Statistical data:**

I. Met data for all agricultural areas stations regarding maximum temperature, minimum temperature, relative humidity, rainfall and sunshine hours
II. 10 daily water availability irrigation data
III. Yield data

a) **Image data:**

NDVI data for last fifteen years:
- NOAA (1992-1998)
- SPOT VGT (1998 onward)

(3) Although basic statistics was covered but less time was given to R-statistical software for regression. Instead of teaching and working on R-software, Statistica and SPSS software should be used for regression with more time on practical exercises.

(4) Teaching sequence may further be enhanced and rescheduled to cover statistical aspect in the start to make it more beneficial.

(5) Lectures on Area Frame Sampling were almost to get more information from participants regarding Area Frame Sampling System in Pakistan.

(6) Lack of availability of all licensed software to individual participants.

(7) No accessibility to the computer labs after office hours in the evening.

(8) Less time for practical work for development of modeling and that too with incomplete statiscal input data in Pakistani context.

(9) All all presentations were delivered in the lecture room.

7. Recommendations:

(1) All input data for last fifteen years required to be used in training may be collected from the concerned departments of Pak Met Department, provincial irrigation departments and concerned ministry and compiled prior to move for the next training session. The said data is reproduced as under:

a) Statistical data:

The statistical comprises:

I. Met data for all agricultural areas stations regarding maximum temperature, minimum temperature, relative humidity, rainfall and sunshine hours
II. 10 daily water data for last fifteen years
III. Yield data

b) **Image data:**

NDVI data for last fifteen years comprising:

- NOAA (1992-1998)
- SPOT VGT (1998 onward)

(2) Training may be carried out by using the above mentioned input data so that each individual/group is able to develop an operational model in Pakistani context for at least one of the major crops of wheat, cotton, sugarcane and rice. For the purpose the efforts in process may be expedited so that the same is ready before departure of the next group of trainees.

(3) It would be appropriate if basic statistical aspect of the training is taught to the trainees here in Pakistan. In this case, we can get benefit from the expertise of already working yield forecast model expert, Mr. Attila Bussay and the coming area frame and sampling designing expert Mr. W. H. Wigton who is basically a statistician.

(4) Course schedule/contents may be reassessed with regard to the sequence and to follow to get more benefits of the training. In this behalf, statistical portion may be placed at the start.

(5) Statistica and SPSS software may be used instead of R-software for regression and each participant are provided the same to work on.

(6) More time may be devoted to the practical work of developing the models for the major crops in next training as compared to the only four days provided during this training schedule. In this regard, it is suggested that at least seven working days be spared for a group/individual to work on developing operational models for all major crops (wheat, cotton, sugarcane, rice) using Pakistani data. These seven working days should be scheduled in last two

weeks of the training so that two to three days are spared for validation, calibration and corrections if any of these models.

(7) A PC or laptop should be available to at least a group of two participants to work after office hours in the evening. This can be arranged in two ways; either university may allow the participants to work in labs in the evening or a laptop to a group of at least two participants be provided from here.