**Principal Component Analysis (PCA)**

PCA is a dimensionality reduction algorithm that uses orthogonal transformation converting the set of observations into a set of linearly correlated values. It is used to discover patterns in a given data set. In this type of analysis, the first principal component has the highest variance and the other principal components succeed according to their respective variance under the condition that is orthogonal to the preceding components. It is used as a means in exploratory data analysis and for making predictive analysis. When a multivariate is detected as a set of coordinates in a high dimensional data space, PCA ensures that it reduces it to a lower-dimensional picture when viewed from its most informative view point. The number of principal components is less than or equal to the number of original variables. The aim of PCA is to acknowledge the directions along which the variation in data is maximum.

The basic approach for PCA is to (i) Collect and standardize the data Set, (ii) Calculate the eigen values and eigen vectors from the covariance matrix, (iii) Sort eigen values in descending order and choose the eigenvectors (iv) Construct the projection matrix from the selected eigenvectors and (v) Transform the original dataset to obtain a dimensional feature subspace. Covariance Matrix is that represents the variance in data set and covariance among the components. By arranging the eigen vectors based on the eigen values in an decreasing order, the principal components are respectively aligned on the basis of their significance. Covariance matrix helps us to relate two variables based on their positive, negative, or non-existing character of their covariance. Eigen values estimates the variation retained by each principal component. In General, eigen values are maximum for first PC and keeps decreasing for the other principal components. An eigen value > 1 indicates that PCs account for more variance than accounted by one of the original variables in standardized data. This is commonly used as a cutoff point for which PCs are retained. This is applicable only when the given data set is standardized.

XLSTAT is software that deals with data analyses and is highly compatible with Microsoft excel. It primarily focuses on statistics and multivariate analysis. It is highly efficient in visualizing the data as histograms, 3-D Plotting, principal component analysis etc; It is easier to install and is user-friendly in nature. The first principal component is the linear collaboration of x-variables that has maximum variance(from the give set of data) so that it accounts for maximum variance in data.

**PCA in analyzing Drought**

The data collected is standardized. In order to analyze the effect of meteorological, hydrological and agricultural droughts on the yield obtained, PCA is used. Since the data collected has raw factors, the principal components are to be chosen. The factors that affect the yield are meteorological, hydrological and agricultural drought and the area under cultivation. The data collected is from 1966-2012. The Procedure for PCA on given set of data is (i) to Install XLSTAT and enable it to work with excel environment, (ii) Select the three principal components, (iii) Using the PCA option in XLSTAT, calculate the eigen value and vectors for the principal components, (iv) The covariance matrix is calculated by the XLSTAT Software (v) Plot the 3-D Scatter graph using the "3-D Graph" option.

**RESULTS AND DISCUSSION**

In this analysis, the meteorological, hydrological and agricultural drought indices are the independent variables (They, in turn, depend on so many related parameters) and the crop area and crop yield are dependent variables. The datasets are subjected to a preliminary analysis of temporal variation of all the independent variables with the crop area and crop yield as well as a correlation analysis to understand the interdependence of the parameters involved. Then, PCS is performed among the related parameters towards identifying the influential type of drought that affects the crop area and crop yield. The following sections present the results of the analysis and discuss the sub-components and in-between results.

**Preliminary analysis of data in Arunanadhi Basin**

**Correlation coefficient:-**

**A negative Correlation coefficient indicates that as variable x increases variable y decreases.**

**A positive correlation coefficient indicates that as variable x increases variable y also increases.**

**PADDY CROP**

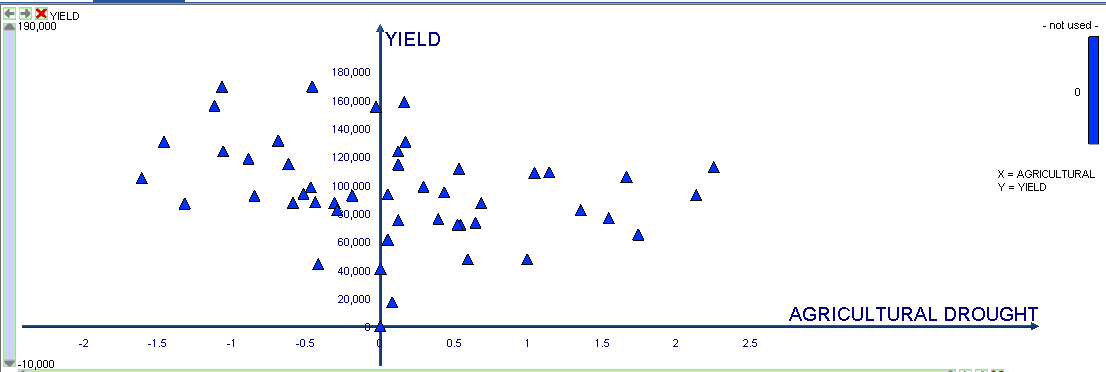
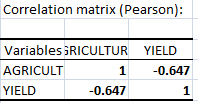
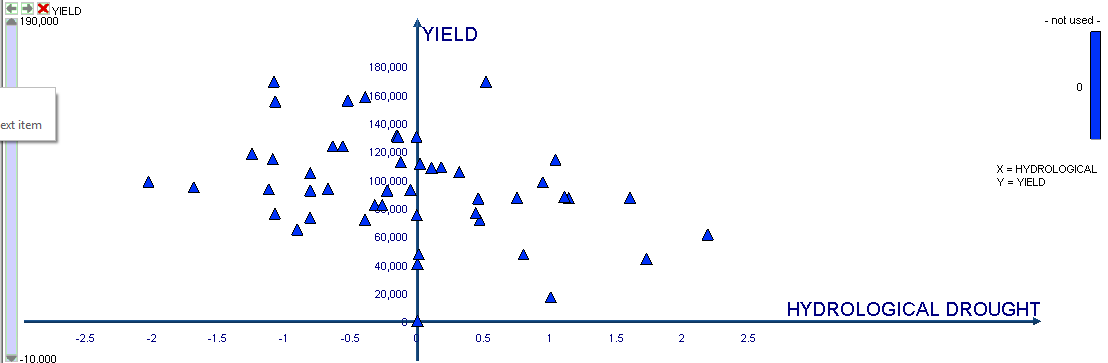
****

fig.agricultural drought vs yield for paddy

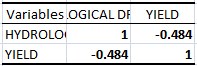
From the graph, we infer that the yield is maximum when the drought parameter is minimum.



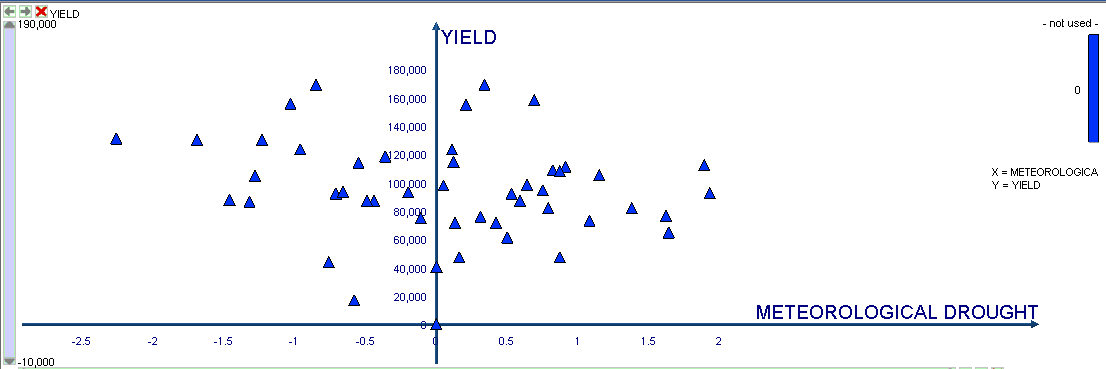
the correlatio coefficient(-0.647) indicates that as agricultural drought increases yield decreases.

****

From the above graph, we infer that yield is maximum when the drought parameter value decreases.



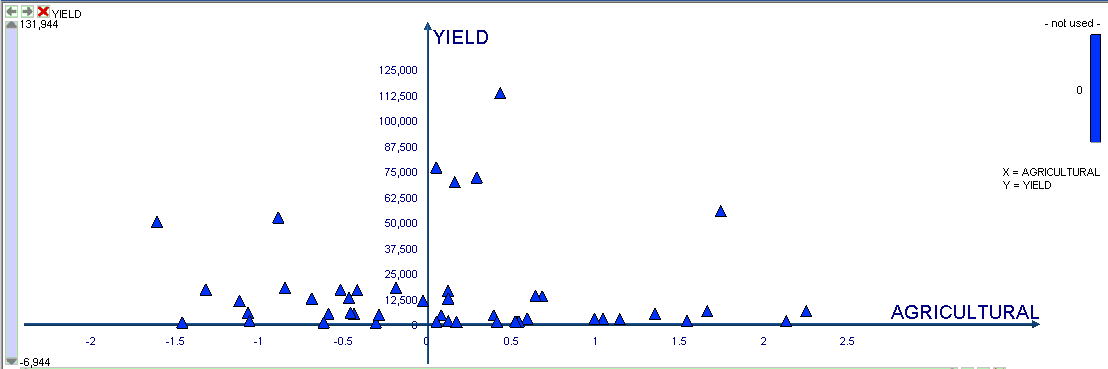
The correlation Coefficient (-0.484) indicates that as hydrological drought value increases yield value decreases.





The correlation coefficient(-0.607) indicates that as drought increases, yield decreases.

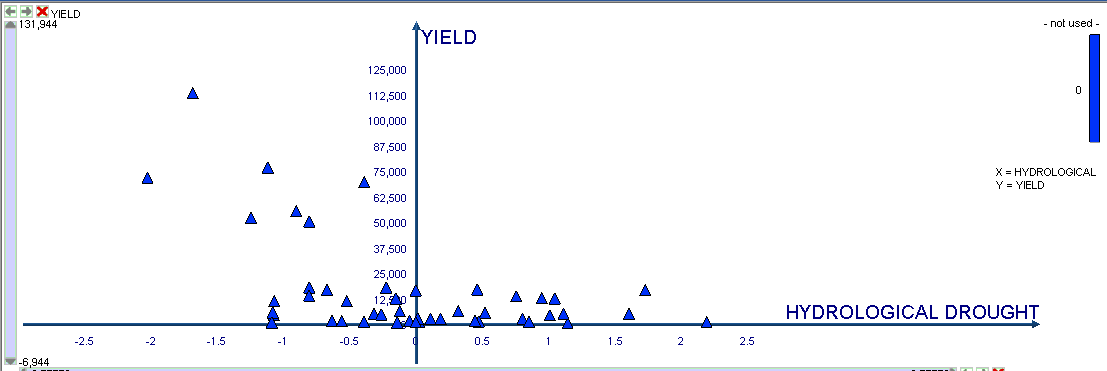
MAIZE



From the above graph we can infer that yield of maize crop decreases with increase in drought values.



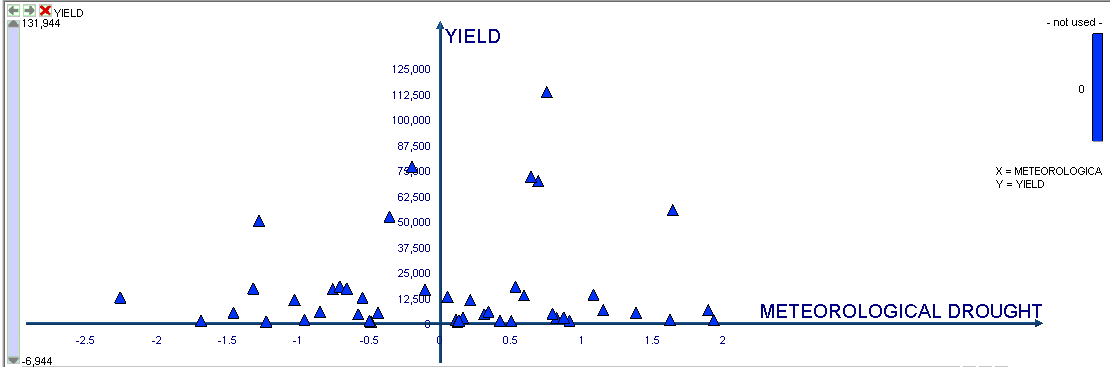
The correlation coefficient(-0.060) indicates that yield decreases with increase in agricultural drought parameter.



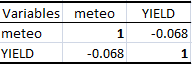
From the above graph we infer that yield increases with decrease in hydrological drought values.



The correlation coefficient(-0.524) indicates that yield of maize crop decreases with increase in hydrological drought values.

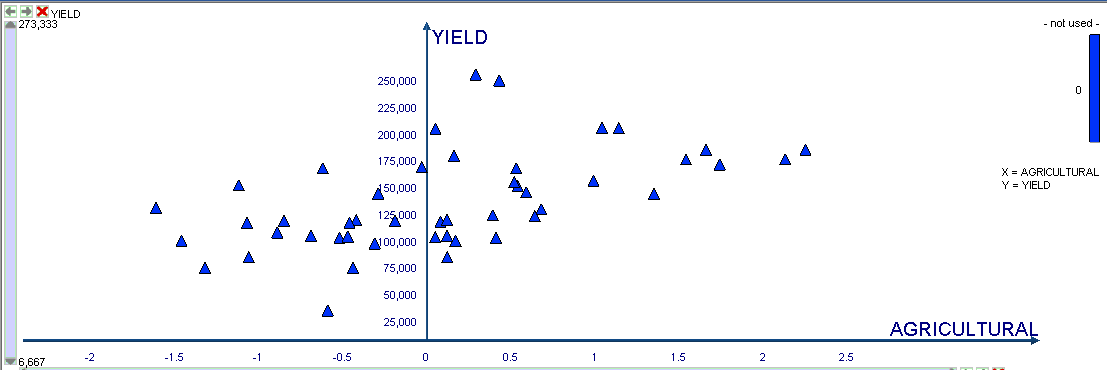


From the above graph we infer that yield increases with decrease in meteorological drought values.



The correlation coefficient(-0.068) indicates that the yield of maize crop increases with decrease in meteorological drought.

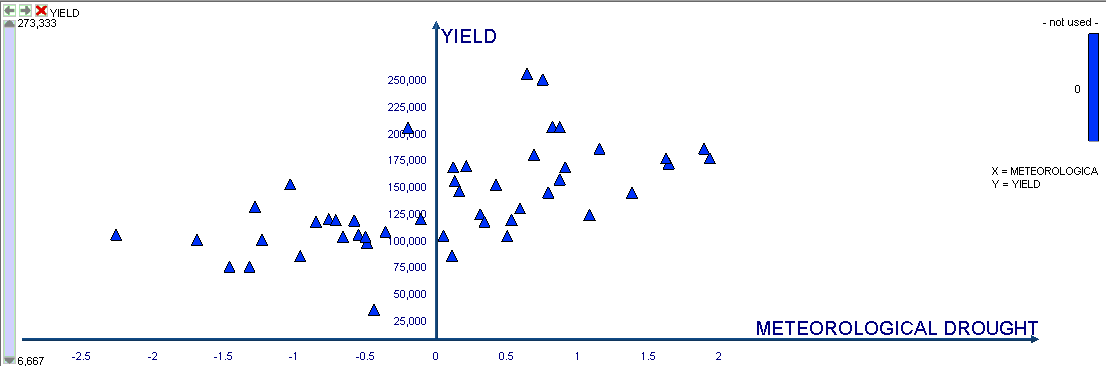
CEREALS



From the above graph we can infer that the yield increases with decrease in agricultural drought values.



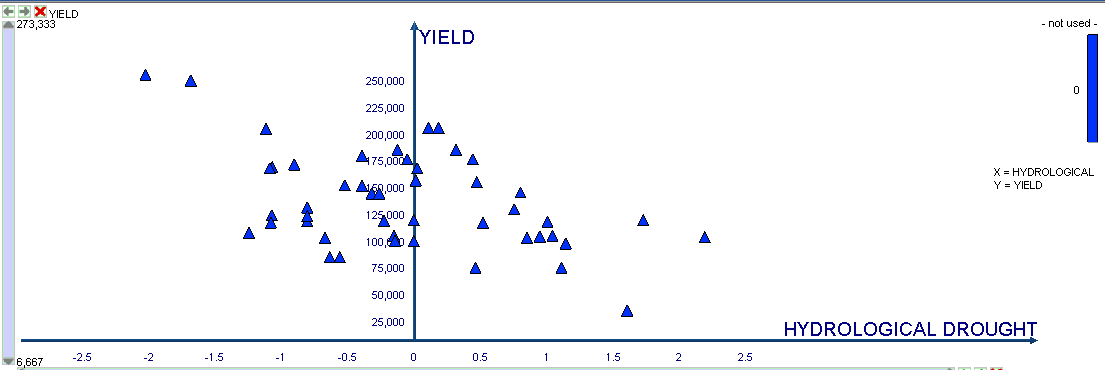
The correlation coefficient(-0.657) indicates that the yield increases with decrease in agricultural drought values.



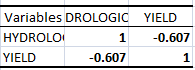
From the above graph, we can infer that yield increases with decrease in meteorological drought values.



The correlation coeffcient(-0.484) indicates that yield increases with decrease in meteorological drought values.



From the above graph we can infer that yield increases with decrease in hydrological drought values.



The correlation coefficient(-0.607) the yield increases with decrease in hydrological drought values.

Calculating PCA(1):-

* Z¹ = Φ¹¹X¹ + Φ²¹X²

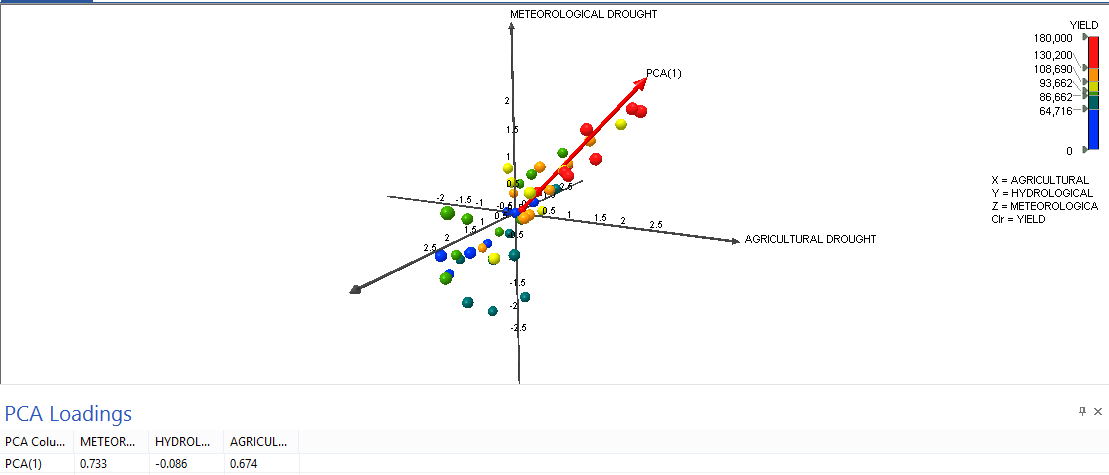
Where Z¹ refers to First Principal Component

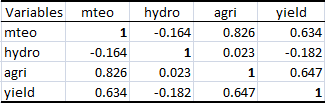
Φ refers to the first two columns present in the transpose of the eigen vectors of the correlation matrix.

X refers to the input data set.

PRINCIPAL COMPONENT ANALYSIS

PCA for paddy crop



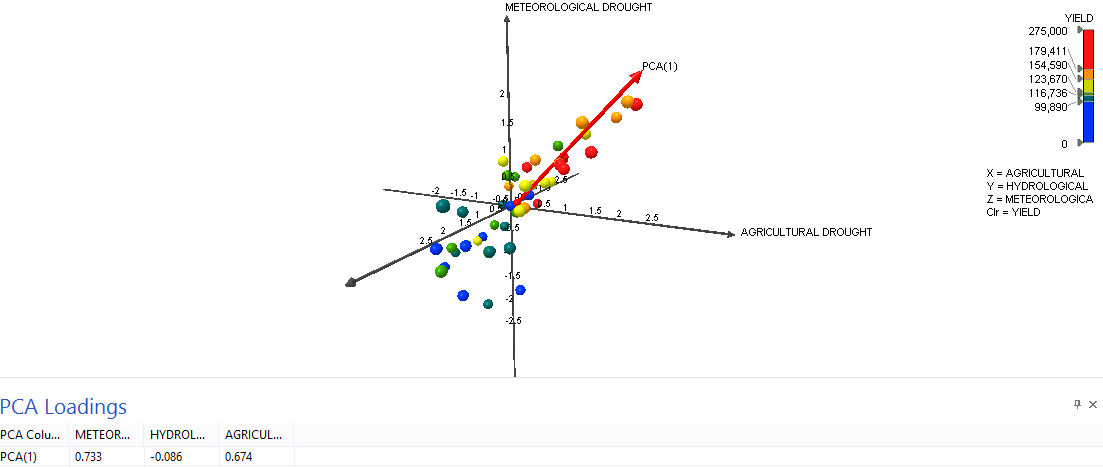


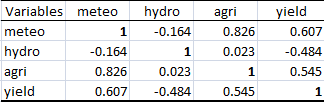
From the PCA values and the correlation matrix we infer that Hydrological drought is highly negatively correlated with the yield. Hence the Arrow in the above graph indicates that yield is minimum when hydrological drought value increases.

C:\Users\USER\Desktop\pca garaph\area vs yield paddy.png

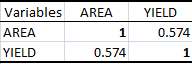
From the above correlation matrix we infer that yield and area are directly proportional to each other and hence they both are linearly dependent on the three principal components:-Agricultural, Meteorological and Hydrological Drought.

PCA for cereals



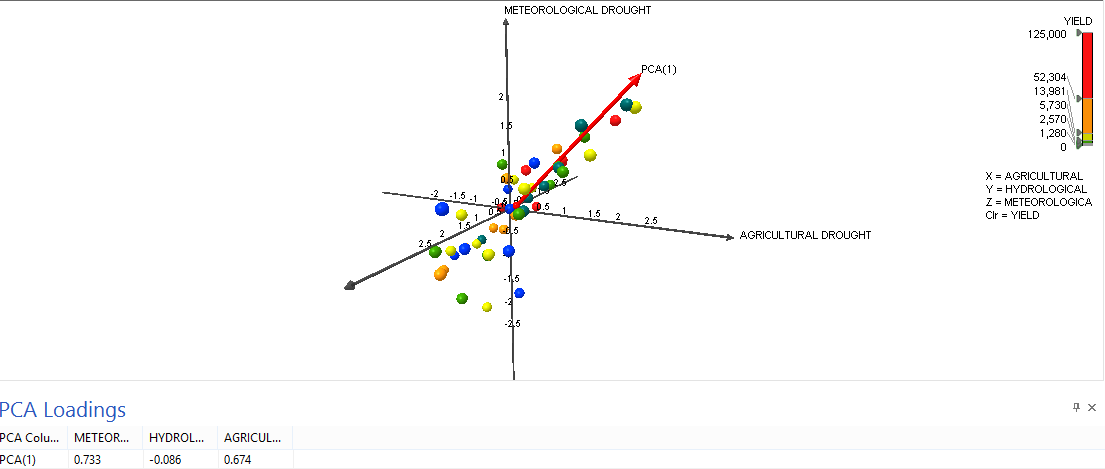


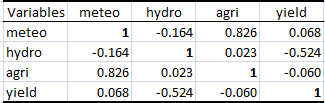
From the PCA values and the correlation matrix we infer that Hydrological drought is highly negatively correlated with the yield. Hence the Arrow in the above graph indicates that yield is minimum when hydrological drought value increases.



From the above correlation matrix we infer that yield and area are directly proportional to each other and hence they both are linearly dependent on the three principal components:-Agricultural, Meteorological and Hydrological Drought.

PCA for maize





From the PCA values and the correlation matrix we infer that Hydrological drought is highly negatively correlated with the yield. Hence the Arrow in the above graph indicates that yield is minimum when hydrological drought value increases.

C:\Users\USER\Desktop\pca garaph\area vs yield maize.png

From the above correlation matrix we infer that yield and area are directly proportional to each other and hence they both are linearly dependent on the three principal components:-Agricultural, Meteorological and Hydrological Drought.

**(Add a few graphs of (Set 1) met drought vs crop yield, hyd drought vs.crop yield and agr.drt vs.crop yield and (set 2) correlation graphs (or values) of these three sets…**

**(See to that the legend of which axis represents which type of drought could be uniform for all three crops)**

**(The red arrow direction / variation with respect to other two crop graphs are qualtified in some way and bring to discussion also)**

**(The vacant space on the left side of graph may be cut, make the plots bigger so that the graph / plot should be “LEGIBLE (readable)”)**

**(If possible, the colour graph may be converted in the form of different shades / patterns, as colour graphs are not adopted by many journals)**

**(Since the PCA table looks the same for all three crops, some parameters may be added to this tabulation so that we try to present the essential variations among the results of three crops)**

**(Also, PCA loading table may be typed in MS word format )**

**CONCLUSIONS**

**(A tentative, draft conclusion paragraph may be written, that can be corrected / modified by me)**