E

UPOV

INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS
GENEVA

TECHNICAL WORKING PARTY ON AUTOMATION AND COMPUTER PROGRAMS

Twenty-Seventh Session
Alexandria, Virginia, United States of America
June 16 to 19, 2009

DESCRIPTION OF MODULES THAT MAKE UP THE DUSTNT SOFTWARE PACKAGE

Document prepared by experts from the United Kingdom

DESCRIPTION OF THE MODULES THAT MAKE UP THE DUSTNT SOFTWARE PACKAGE

1.   DUSTNT is Windows software that helps to manage and analyze data from DUS spaced-plant variety trials for cross-pollinated varieties - such as herbage and certain vegetable crops. Data are recorded on the DUS trials several times during the growing season. On each occasion, a different set of characteristics is measured on all the plants. The data are stored in data files. The DUSTNT software comprises a series of modules or programs linked by data files. DUSTNT allows the user to accumulate data over the growing year, and summaries and analyze them. As this happens, the nature of the data stored in the resulting data files changes, e.g. from measurements on individual plants to plot means to within-year variety means to over-year variety means. Different types of data are stored in different types of data file.

2.   The DUSTNT software allows many kinds of analyses. They include COY analyses and a wide range of multivariate analyses. These analyses reflect the history of the development of statistical approaches in DUS testing: a history driven by the need for valid approaches and efficient use of data. This history is documented in the TWC working papers. Once a module had been written to perform a particular analysis, the value of that analysis in DUS testing could be investigated. As a result the DUSTNT software acts as a toolkit enabling many different analyses, some of which are in regular use for DUS testing.

3.   Below is a list of the modules and their functions. Symbol # indicates modules that may form part of a COY analysis, including the long-term-COYD criterion and the COYD criterion applied to unbalanced data, e.g. when trials are conducted using the Cyclic Controls approach. Symbol * indicates modules used in the UK for analyses subsequent to COY, e.g. to review characteristics for importance, to investigate problem variety pairs, and to derive variety descriptions. The module list is followed by a table showing which other modules may need to be run before a module can be used.

Module	Function
# **LOGR9**	Converts datalogger file to an R file (plant data) for input to module SUMM9
# **RMRG9**	Merges a number of R files (plant data) on different batches of characteristics from the same trial into another R file for input to module SUMM9
# **SUMM9**	Reads plant data from R file (plant data), checks it, summarizes it, optionally produces histograms and calculates characteristic transformations, and calculates plot means and within plot standard deviations which are output to a D file (plot data).
# **TRAN9**	Operates on plot means from a D or E file (plot data) to either discard characteristics or create new ones by combining measured characteristics in various ways on a plot basis and output to a new D or E file.
# **ANAL9**	Carries out univariate randomized block analyses of variance on plot means for all characteristics in a D or E file (plot data) and produces an M file (variety means) and a J file (variety within plot standard deviations). It can also operate on an E file of variety means from several years created by FIND9 to produce an M file of over-year means.

#CHOS9   To combine data on the same characteristics from one or more D files (plot data) into another D file omitting specified varieties, e.g. to combine data from several maturity groups of a trial into a single file or to omit several varieties from a single D file.

#DMRG9   Merges a number of D files (plot data) on different batches of characteristics from the same trial into another D file for input to module ANAL9.

#TEST9   Compares nominated candidate varieties with every other variety on all measured or derived characteristics using t-tests on the means from an M file, i.e. a within-year test, or as part of the long-term-COYD criterion.

#DUST9   For within-year comparisons of varieties (or as part of the long-term-COYD criterion) this module does an analysis of the characteristics to answer:-

–   What are the numbers of varieties each characteristic separates?

–   How similar are a pair of varieties based on the measured characteristics?

–   Are there varieties which cannot be separated by the measured characteristics?

–   Are there varieties which can only be separated by one characteristic i.e. an essential characteristic?

–   What is the minimum characteristic set to achieve the same separations as obtained by all the characteristics together?

COVR9   To do a covariance analysis on data held in an E file (plot data) using data held in another E file as a covariate.

*UNIF9   Provides a within-year uniformity test of candidate varieties against selected control varieties on the basis of the following criteria: (mean standard deviation + t value * standard deviation of the standard of control varieties). This also may be done after regression and adjustment of within plot standard deviations on variety means by module REGR9.

#OMIT9   Omits defined varieties and characteristics from an M file (variety means) or J file (variety within plot standard deviations) and writes the remaining data to a second file of the same type.

#GRUP9   Combines the M files (variety means) from a number of trials in one year into a single M file omitting LSD's. It is expected that the new M file will be used in conjunction with FITC9 or FITR9 to produce variety descriptions over years or centers.

PLOT9   Extracts data for several varieties on a single characteristic from an R file (plant data) into a file suitable for input into a spreadsheet, e.g. for graphing.

#YEAR9   Replaces SE's, LSD's and DF's in a one year M file (variety means) by values of the same statistics derived from an over-years file produced by FITR9 or FITC9. This enables one year variety means to be evaluated using over-year statistics.

ORDR9   Put the varieties in a D or E file (plot data) into variety sample number order.

*REGR9   Regresses single year within plot standard deviations (or log standard deviations) on variety means, provides a scatter plot of the results, and adjusts the within plot standard deviations for the regression.

# **TVAL9**	Performs over-year distinctness tests using the UPOV COYD criterion and reports detailed results of comparisons between candidates and all other varieties
# **TSUM9**	Performs over-year distinctness tests using the UPOV COYD criterion and provides a condensed summary of overall distinctness results achieved between candidates and all other varieties according to the COYD criterion.
# **TVRP9**	Performs over-year distinctness tests using the UPOV COYD criterion. Reports detailed results of those comparisons between candidates and other varieties which are not distinct according to the COYD criterion, and provides a condensed summary of overall distinctness results achieved according to the COYD criterion.
* **DUS9/INTR9**	DUS9 is an interactive program, run in MSDOS, to enable specified pairs of varieties to be compared within years or over years.
	INTR9 is the Windows version of DUS9.
# **COYU9**	Performs over-year uniformity tests using the UPOV COYU criterion.
# **FIND9**	Extracts variety means on defined varieties and characteristics from M files from a number of years for a trial and puts into an E file in which the replicates correspond to the years of the M files. The E file so formed may be submitted to ANAL9, SSQR9 or FITC9 for further analysis.
DIAG9	Provides plots of variety means on pairs of characteristics.
* **MULT9**	Analysis to determine the characteristics that are likely to be most important in separating varieties. Calculates multiple correlations between characteristics to identify characteristics which are least dependent on all others and hence are likely to be the most important in separating varieties.
* **SSQR9**	Calculates within and between varieties co-variance and correlation matrices, having first removed the effect of blocks from the plot mean data or the effect of years from mean year data.
PCMP9	Performs a principal component analysis on a defined co-variance or correlation matrix. In the DUS context it is used to evaluate important features in the relationship between the characteristics independently of the varieties grown in a particular trial, and so is used in conjunction with the within variety variation.
* **DIST9**	Calculates Malahanobis multivariate generalized distances (D^2) between variety pairs.
CANV9	Performs the multivariate canonical variate analysis to identify functions of characteristics that maximize the variation between varieties relative to the variation within varieties.
PCRD9	Calculates coordinates of variety means on principal coordinate axes derived from an array of distances between all variety pairs. These distances can be CRL values, D values or similarity coefficients.
CLUS9	Performs multivariate cluster analysis using a choice of methods
GOWR9	Calculates dissimilarities between taxonomic units using Gower's general coefficient of similarity

CORR9   Calculates correlation coefficients and produces plots between variety means on defined characteristics in two years for a particular trial using data stored in an E file in which the replicates are the years.

CORL9   Calculates correlation coefficients (r) between characteristics using variety means in an M file and optionally outputs distances between characteristics to an S file and the between variety variance/covariance matrix to a B file.

[#] **MNML9**   Provides, on the basis of conditional F ratios, an ordering of characteristics commencing with the one which is most discriminating followed by the one which is most discriminating given that the first characteristic is already in use. This process is repeated until all characteristics are ordered. This can be done either on within-year data or over-year data.

[*] **MOST9**   Determines the most similar variety to each of a specified set of varieties based on data extracted from M files over a given number of years. The most similar variety may be defined as the one with the smallest Mahalanobis distance with respect to a specified variety. Alternatively, the most similar variety may be defined as the one with the smallest maximum t-value on all characteristics.

[*] **PBLM9**   Examines the multivariate distances between 2 varieties on all pairs of characteristics to enable identification of possible characteristic pair combinations important to separating the 2 varieties.

MAHL9   Provide a profile of Mahalanobis' distances over all characteristics for particular variety pairs. This is constructed as follows: The characteristic which separates the variety pair by the largest D^2 value is first chosen. Then the characteristic which increases this D^2 value by the greatest amount when combined with the first characteristic is added. Further characteristics are added using the same procedure until all characteristics have been accounted for. At each stage a test of significance of the characteristics remaining is carried out using a test given by Rao.

[#] **FITC9**   Produces adjusted means over years or sites using a fitted constants analysis for a number of characteristics. Missing values are permitted. Data are read from an E file produced by FIND9 or by RETR9. The fitted constants analysis is calculated using matrix inversion.

[*] **VDES9**   Provides variety descriptions on the basis of over-year variety means. A description consists of the class numbers within a defined range for each variety in a species on a set of characteristics. Class numbers can be any set of n equally spaced numbers viz 1, 3, 5, 7, 9. A class number is obtained either by dividing the range of variety means on a species into n equal sections or by delineating class boundaries by means of n-1 'seeded' varieties. The value of n is defined for each characteristic separately. The variety descriptions are output to an M file.

[#] **FITR9**   Produces adjusted means over years or sites using a fitted constants analysis for a number of characteristics. Missing values are permitted. Data are read from an E file produced by FIND9 or by RETR9. The fitted constants analysis is calculated using the FCSUB routine provided by SASS.

SNTV9

Modified joint regression analysis is used to determine variety sensitivities from data extracted from M files. These variety sensitivities are the slopes of linear regression lines of individual variety means against overall variety means in a number of environments (sites or years).

* SAME9

Detects varieties with the same variety description using the M file produced by module VDES9.

COPH9

Reads distances between all pairs of varieties in a set from an 'S' file and corresponding distances derived by CLUS9 from a dendrogram of the same set of varieties, and from them provides a correlation coefficient referred to as a cophenetic correlation coefficient.

 # VMRG9

Merges a number of R files (plant data) containing data on same characteristics for different batches of varieties (e.g. different maturity groups) from the same trial into another R file. The merged R file may then be further merged with other R files by RMRG9 to produce a master R file containing plant data for a complete trial.

 # RETR9

Used when trials are conducted using the Cyclic Controls approach where not all reference varieties are planted each year but instead are planted cyclically two years out of three. Retrieves data for the purpose of undertaking an unbalanced COYD analysis by FITC9 or MJRA9. Data are written to an E file which is submitted to FITC9 or MJRA9 for further analysis.

 # UNEQ9

Used when trials are conducted using the Cyclic Controls approach. Provides a condensed summary of overall distinctness results achieved by specified candidate varieties over a several year period using the COYD criterion applied to unbalanced data. Also, provides detailed reports for specific variety pair comparisons. The choice of variety pairs depends on whether interim or final reports are being considered.

 # MJRA9

Produces adjusted means over years or sites using a modified joint regression analysis (MJRA) for a number of characteristics. In this analysis regression lines are fitted to variety year means versus variety means over all years i.e. one line per year. Missing values in the data input file are permitted. Data are read from an E file produced by RETR9.

* CHAR9

Carries out a characteristic review following analysis by MJRA9 or FITC9. This characteristic review is similar to that undertaken by DUST9 on within-year data but is done on over-year data. Thus the number of varieties separated by each characteristic and the minimal characteristic set are produced, and all essential characteristics are identified.

 # CYCL9

Used when trials are conducted using the Cyclic Controls approach. Combines the functions of RETR9, MJRA9, FITC9 and UNEQ9 by retrieving data on specified varieties, performing a modified joint regression or fitted constants analysis on the retrieved data and presenting the distinctness status of candidate varieties and results of comparisons between specified variety pairs.

 # INCB9

Carries out the analysis of a trial with a resolvable incomplete block design. Data are held in a D (or E) file (plot data) obtained using SUMM9. The incomplete block codes are taken from an H file (header data). The adjusted variety means and approximate LSD's are output to an M file (variety means).

# **CONV9**	Prepares files for upload to data logger for an incomplete block design trial prior to data collection.
# **MMRG9**	To combine 2 or more M files (variety means) relating to the same trial. The merged M files do not need to contain identical sets of varieties as MMRG9 will produce a file with only those varieties common to all the input M files.
# **COMB9**	Combines the functions of FIND9 and FITP9 by retrieving data from several single year M files (variety means) and deriving from them a new M file containing means for varieties with complete data over a specified set of years and with SE's and LSD's based on all years and all varieties. The output M file may be further analyzed by TEST9 or DUST9 as part of the long-term-COYD criterion.
# **FITP9**	Reads an E file, normally produced by FIND9 or RETR9, and derives from it an M file containing means for varieties with complete data over a specified set of years and with SE's and LSD's based on all years and all varieties. The output M file may be further analyzed by TEST9 or DUST9 as part of the long-term-COYD criterion.

	Other modules which may need to be run beforehand																		
Module	ANAL9 #	COMB9 #	CORL9	DIST9 *	DMRG9 #	DUST9 #	FIND9 #	FITC9 #	FITP9 #	FITR9 #	GOWR9	LOGR9 #	MJRA9 #	REGR9 *	RETR9 #	RMRG9 #	SSQR9 *	SUMM9 #	VDES9 *
ANAL9					√		√											√	
CANV9	√																√		
CHAR9						√							√						
CHOS9																		√	
CLUS9	√		√	√							√						√		
COMB9	√																		
CONV9																			
COPH9	√			√		√	√				√						√		
CORL9	√																		
CORR9						√													
COVR9																		√	
COYU9	√																		
CYCL9	√																		
DIAG9	√																		
DIST9	√																√		
DMRG9																		√	
DUS9/INTR9	√																		
DUST9	√	√						√											
FIND9	√																		
FITC9							√								√				
FITP9							√								√				
FITR9							√								√				
GOWR9																			
GRUP9	√																		
INCB9																		√	

Module	ANAL9#	COMB9#	CORL9	DIST9*	DMRG9#	DUST9#	FIND9#	FITC9	FITP9#	FITR9#	GOWR9	LOGR9#	MJRA9#	REGR9*	RETR9#	RMRG9#	SSQR9*	SUMM9#	VDES9*
														Other modules which may need to be run beforehand					
LOGR9																			
MAHL9	√						√										√		
MJRA9	√														√				
MMRG9	√																		
MNML9							√											√	
MOST9	√			√			√										√		
MULT9																	√		
OMIT9	√																		
ORDR9																	√		
PBLM9	√						√										√		
PCMP9	√		√			√					√						√		
PCRD9	√		√	√													√		
PLOT9												√							
REGR9	√																		
RETR9	√																		
RMRG9																			
SAME9																			√
SNTV9							√								√				
SSQR9							√												
SUMM9												√			√				
TEST9	√	√						√											
TRAN9																		√	
TSUM9	√																		
TVAL9	√																		
TVRP9	√																		
UNEQ9	√												√		√				
UNIF9	√														√				
VDES9								√		√									
VMRG9												√							
YEAR9	√						√			√									

[End of document]