Volume 3

ICASA Tools

Volume 3

DSSAT v4.5
ICASA Tools

Editors:

James W. Jones
Gerrit Hoogenboom
Paul W. Wilkens
Cheryl H. Porter
Gordon Y. Tsuji

International Consortium for Agricultural Systems Applications
University of Hawaii

IBSNAT, The International Benchmark Sites Network for Agrotechnology Transfer, was a network consisting of the contractor (University of Hawaii), its subcontractors and many global collaborators. Together they created a network of national, regional, and international agricultural research for the transfer of agrotechnology among global partners in both developed and developing countries.

From 1982 to 1987, IBSNAT was a program of the U.S. Agency for International Development under a cost-reimbursement Contract, No. DAN-4054-C-00-2071-00, with the University of Hawaii. From 1987 to 1993, the contract was replaced with a Cooperative Agreement, No. DAN- 4054-A-00-7081-00, between the University of Hawaii and USAID.

In 1994, the International Consortium for Agricultural Systems Application or ICASA was established as a non-profit corporation in Honolulu, Hawaii. ICASA is governed by a Board of Directors consisting of former members of the IBSNAT network and international systems scientists. The ICASA network oversees the development of systems tools, including DSSAT.

Correct citation: J. W. Jones, G. Hoogenboom, P.W. Wilkens, C.H. Porter, and G.Y. Tsuji (Editors). 2010. Decision Support System for Agrotechnology Transfer Version 4.5. Volume 3. DSSAT v4.5: ICASA Tools. University of Hawaii, Honolulu, HI

LIBRARY OF CONGRESS 94-19296
ISBN 1-886684-08-1

Volume 3

Table of Contents

DSSAT V4.5

Volume 3
Chapter 1

Genetic Parameter Estimation Tool (GenCalc)

L.A. Hunt

University of Guelph, Guelph, Ontario, Canada
International Consortium for Agricultural Systems Applications

Note:
To change the product logo for your own print manual or PDF, click "Tools > Manual Designer" and modify the print manual template.

GenSelect Help

Title page 1

Use this page to introduce the product

by

This is "Title Page 1" - you may use this page to introduce your product, show title, author, copyright, company logos, etc.

This page intentionally starts on an odd page, so that it is on the right half of an open book from the readers point of view. This is the reason why the previous page was blank (the previous page is the back side of the cover)

GenSelect Help

Printed: July 2006 in (whereever you are located)

Publisher

...enter name...

Managing Editor

...enter name...

Technical Editors

...enter name...
...enter name...

Cover Designer

...enter name...

Team Coordinator

...enter name...

Production

...enter name...

Special thanks to:

All the people who contributed to this document, to mum and dad and grandpa, to my sisters and brothers and mothers in law, to our secretary Kathrin, to the graphic artist who created this great product logo on the cover page (sorry, don't remember your name at the moment but you did a great work), to the pizza service down the street (your daily Capricciosas saved our lives), to the copy shop where this document will be duplicated, and and and...

Last not least, we want to thank EC Software who wrote this great help tool called HELP & MANUAL which printed this document.

Table of Contents

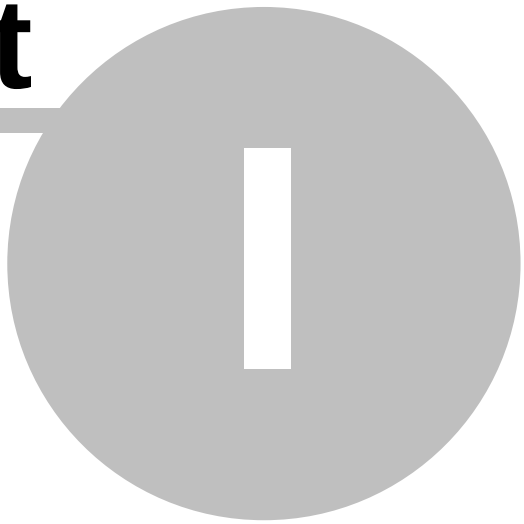
Foreword

This is just another title page
placed between table of contents
and topics

Top Level Intro

This page is printed before a new top-level chapter starts

Part



1 Introduction

1.1 Welcome to GenSelect/GenCalc

GENSELECT is software that makes it possible to select and organize experiment/treatment combinations for use when estimating coefficients for a specific cultivar using the software 'GENCALC'. The selected combinations are listed in a file, the name of which is constructed from the cultivar name, the crop code (e.g. SB for soybean), and a final 'C' (e.g. BRAGG.SBC).

GENCALC2 is software that makes use of a crop model to determine cultivar trait coefficients from data reported for an array of experiments.

1.2 GenSelect

A short tutorial follows on using the Genotype Selector software in DSSAT.

1.2.1 Choice of Exisiting 'LIST' file

From the main screen, press 'Next' to proceed to the first working screen. Choose the desired crop and examine the list of experiment/treatment combination files that are already present. If there is a file for the cultivar you wish to examine, select the file by highlighting it and then press the 'View the file' bar to determine if the experiment/treatment combinations you wish to use are included. If so, press the 'Use the existing selected file' bar and then press 'Next' to proceed to GENCALC per se. If not, press 'Back' and proceed to make a new 'List' file.

1.2.2 Create a new 'LIST' file

From the main screen, press 'Next' to proceed to the first working screen. Choose the desired crop and press 'Next' and a list of cultivars in experimental files in the main experiment file directory, as indicated in the
right hand box, will appear. If you wish to explore other directories, activate the 'Scan additional ...' menu item and select the desired directories. Highlight the cultivar you wish to work with, press 'Next', and then select the experiment/treatment combinations to be used, remembering that it is best to first work with the non-stressed treatments, and press 'Next' to proceed to save the new file, which is named with the chosen cultivar name.

1.3 GenCalc

GENCALC2 is software that makes use of a crop model to determine cultivar trait coefficients from data reported for an array of experiments. The model must use CULTIVAR and ECOTYPE files as specified for the DSSAT models, and must generate an EVALUATE.OUT file which conforms to the standards set up for DSSAT. Additionally, the model must make use of a configuration file similar to the DSSATPRO file of DSSAT, and be launchable with a command line argument that contains information on the experiment/ treatment combinations to be used. For the DSSAT CSM model, the command line could have (depending on general configuration) the following form:

C:\DSSAT4\TOOLS\GENCALC2 BRAGG.SBC DV4

In which BRAGG.SBC is the name of a file containing a list of experiment/treatment combinations to be used. Files of this type, which can be created using the GENCALC selector program, must have the form shown below:

$BATCH(CULTIVAR):SBIB0001 BRAGG

```
@FILEX                          TRNO  RN
C:\DSSAT4\GCWork\UFGA7801.SBX          1   0
```

C:\DSSAT4\GCWork\UFGA7801.SBX 2 0

The DV4 indicates that the program is to run using the DSSAT configuration file (Dssatpro.fle).

The name GENCALC stands for Genotype Coefficient Calculator, and the digit '2' indicates that the program is an update of the original calculator that was included in the DSSAT 3.5 package, and described in the following publication:

Hunt,L.A.,S.Pararajaysingham,J.W.Jones,G.Hoogenboom, D.T.Imamura,and R.M.Ogoshi. 1993 GENCALC:Software to facilitate the use of crop models for analyzing field experiments. Agronomy J. 85(5):1090-1094

1.3.1 Mode of Operation - Cultivar Cefficiients

The program works by running the model with a number of different sets of cultivar coefficients. The outputs from the model runs are examined, and the coefficient set which results in the least difference (evaluated using the Root Mean Square Error) between simulated and measured values over all runs is selected.

The initial set of cultivar coefficients is usually obtained from the cultivar file using the cultivar identifier. But if no entry with the identifier is present in the file, they are obtained using a default cultivar descriptor. This default cultivar must be chosen from among those included in a list presented in a selector box on the computer screen. This list box is populated with all entries in the cultivar file which have a code beginning with 99. If there is only one default entry with a '99' code, no list box is presented and the unique entry is used to provide the initial set of coefficients.

The sequence of model runs is defined in a so-called RULES file. In this file, the aspect that is to be used to determine error (the 'Target' trait) is defined along with the coefficients that are to be varied, the amount by which each coefficient is to be varied, and the number of simulations to be performed. The rules file has the form shown below:

TARGET..... STEP,COEFF,LOOPS
ADAP FLOWERING DAP 10.0,CSDL,3 10.0,EM-FL,3

in which the standard abbreviation for the target aspect is given along with a brief description, followed by information on the coefficient(s) to be changed.

The coefficients are defined using the abbreviations from the cultivar coefficient file. Each abbreviation is preceded by a 'step' value, and followed by a 'loop' number. The step value defines the amount that the coefficient will be changed, given as a percentage of the value in the coefficients file; and the step value defines the number of simulations that will be performed within this range of values.

Thus, if the step is entered as 10.0, the first simulation would be run with a coefficient 5% less than the value in the coefficients file, the last run with a value 5% greater. And if the loop number was defined as 3, then an additional simulation would be run with a coefficient equal to the original file value. Naturally, if the initial coefficient value was zero, no change would be effected. For such zero value coefficients, the loop number should be set to 1 to avoid duplicate 'zero value' runs. Alternatively, a non-zero coefficient value could be entered in the cultivar file.

Depending on the entries in the rules file, the model will run for one 'target' trait, then for a second, and so on until the coefficient changes for the final trait have been run. The coefficients which produce the least error for the first target are used during runs with the second target, and so on

throughout the rules package.

1.3.2 Mode of Operation - Ecotype Coefficients

The cultivar optimization can be conducted with a number of different sets of ecotype coefficients, so permitting an optimization for one or more of the ecotype coefficients. To do this, additional sets of ecotype coefficients should be added to the 'GENCALC' ecotype file before starting a calibration run.

To edit the GENCALC ecotype file, activate the 'Files', 'Coefficients','GENCALC Ecotype File' menu, duplicate (as many times as desired) the single line of coefficients that will be present in the file, change the values of the coefficients that are to be examined, and save the file.

1.3.3 Changing Configuration, Rules, Exp/Treatment List Files

The configuration, rules, and experiment/treatment list files can be accessed and edited by activating the appropriate item in the 'Files' menu list.

In 'configuration', the editor, graphics program, and model can be changed; in 'rules', existing rules can be deleted, changed or commented out (with a ! in the first position), or new rules added; and in the 'list' file, existing treatments can be eliminated or commented out (with a ! in the first position), or new ones in which the same cultivar is present can be added. After accessing any of these files, even if no changes are made, the run number is refreshed.
.

1.3.4 Setting the Rules

General sets of rules are included in the Rules file and can be applied 'as is'. However, results from calibrating the timing of later stages in the growth cycle are not truly meaningful until the earlier stages have been well calibrated. Equally, calibration of growth and yield aspects is not meaningful until phenology is well simulated. Rules dealing with the later growth stages, and with growth and yield aspects, could thus be commented out at the start of a calibration session.

Further, the span specified in the general rules set for changing the coefficient values is, of necessity, an intermediate, general value. The span could thus usefully be increased for early runs in a calibration session, reduced for later runs.

1.3.5 Checking the Datasets

Before proceeding to calibration, a 'check' run should be made to ensure that the model will run with all experiment/ treatment combinations. This could be done by running from the model interface but, when a lot of experiment/treatment combinations are to be examined, is perhaps best done by choosing the 'Check' option from the 'Action' toolbar item.

If the model does not run with one of the combinations, a message will appear on the screen and the run will be terminated. Examine the model Error.out file to determine the reason for failure, and correct the problem before proceeding. Alternatively, the 'offending' experiment/ treatment combination could be removed from the list file.

Once the model runs with all combinations, proceed to examine the Evaluation.out file to ensure that the simulated emergence dates either match well with the recorded dates, or appear reasonable. If this was not
the case, check the individual experiments to determine the cause of a delayed emergence, and correct whatever problems were identified before proceeding with calibration.

Finally, check the Overview.out file to ensure that the correct weather data had been used for the individual runs. This could happen if the planting date had been erroneously entered as sometimes happens with crops that are planted in one year, harvested in another.

1.3.6 Running

A calibration run is initiated by activating the 'Action' - 'Run' menu item. Before initiating a comprehensive calibration run, however, users should determine if the time that will be needed for the calibration run fits in with other demands on computer and personal time. This can be done by examining the 'Runs needed' box on the computer screen and then making a quick mental calculation of the total time necessary for the run. The box indicates the number of individual simulation runs that the contents of the list and rules files call for, so that the calculation can be easily made with pre-knowledge of the time necessary for one simulation.

Once initiated, the progress of the calibration run is indicated in a 'Runs completed' box at the base of the screen. However, because the model uses the list file, the number of runs completed is only up-dated when all experiment/treatment combinations in the list file have been completed for one set of coefficient values.

If it is desired to stop a calibration run, activating the 'Cancel' button at the base of the screen will stop the run as soon as control is returned from the model to Gencalc2 (i.e. when all experiment/ treatments in the list file have been completed for the coefficients being examined at the time of activation).

1.3.7 Examining the Results

Gencalc2 produces a number of files showing the results of the succession of simulation runs, together with the coefficients used in the runs. These files can be examined by activating the menu items under 'Files', 'Errors'.

The 'Details' file contains the simulated and measured values for each individual model run, 'Summary' the errors after running all experiment/treatment combinations with one of the required sets of coefficients, sorted so that runs with the smallest errors appear at the top. And finally 'Minima' with the 'best' results for the runs with different target traits, and 'Comparison' with errors for all traits for which data are available (not only those used as targets). This latter file, in which errors using the original coefficients are compared with those using the 'best' coefficients from the calibration run, is probably the most useful in deciding whether to upgrade the cultivar file with the calibration results.

Before proceeding to these 'error' files, however, it is worthwhile examining the 'Log' file to see if Gencalc2 had detected any missing simulated data. If so, there would have been some problem either with the coefficients used for a particular run (e.g. a vernalization requirement that prevented the crop proceeding to flowering and maturity), or with the model code. In both cases, the causes of the problem should be examined in some detail before proceeding further with an errors examination.

The simulation outputs using the final 'best' set of coefficients can be examined by activating the 'Model Outputs' menu item. The Overview gives detail for each run with the 'best' coefficients, whilst the Evaluate gives a summary comparison of measured and simulated values that can be viewed or plotted, as desired. The graphing program in DSSAT V4.0.2 does not differentiate between sub-set of data in the Evaluate.out file, and so it is not easy to gain a visual impression of the gains made through calibration. Inspection of the file, or use of an alternative graphing program, however, can provide can provide an insight into gains and/or losses in precision and accuracy.

When examining results it is important to bear in mind that errors in the timing of later stages in the life cycle may show an increase when the simulation of earlier stages was initially poor. Errors in various growth and yield aspects also may increase when the timing of the various growth stages was originally poorly simulated.

1.3.8 Upgrading Coefficients using Calibration Results

The coefficients with the least error for the final target represent the 'best' overall coefficients for all targets, and if desired (after examination of the outputs) these can be used to replace the existing

coefficients in the cultivar file. This is done by activating the automatic edit option from the 'Files', 'Coefficients','Main cultivar file','Automatic' menu. When this is done, both cultivar and ecotype coefficients are upgraded.

When coefficients are upgraded, the original coefficients are retained in the file, but with a ! in the first column position. To avoid 'cluttering' the files with such saved coefficients, a user should thus occasionally manually edit both the cultivar and ecotype files and delete saved entries from earlier calibration sessions.

Further, at the start of an upgrade, both the cultivar and ecotype files are copied to backup files. These should be used to replace the main files if a power failure or other unusual event whilst upgrading results in damage to the main files. .

1.3.9 Multi-Run Calibration Sessions

It is unlikely that a cultivar will be well calibrated following one run using the standard rules set. Indeed several runs, each following an upgrade of the coefficients and possibly also a change in the rules, will usually be necessary. Such repeated runs can be initiated without exiting Gencalc2. A calibration session thus generally proceeds as follows:

1. Check that the model runs with all experiment/treatments
2 .Run with the standard rules set
3. Examine the log file to determine if there were any problems And then, if all was well:
4. Examine the comparison file, Possibly examine some of the other error/model files
5. Upgrade or manually edit the coefficient files; Edit the rules file if thought necessary
6 Run again. Before attempting to run, however, check the screen box labeled 'Runs needed'. A zero number indicates that nothing has been changed from the previous run, and that there is need for upgrading or editing coefficients, or editing the rules. And then re-examine the comparison file and repeat the subsequent steps until there is little difference between results for runs with the starting and ending coefficients. At this stage it may be useful to edit the rules file and reduce the step size for each of the components in each rule set, and then repeat from item 4..

1.3.10 Log File and Run Problems

All activities during a session are written to a 'log' file which can be examined by activating the Log menu item under Files. All Gencalc2 error messages are written to this file (even though they may first be shown in a message box). A model error will be identified in the log file, but the actual model error file must be accessed through the Files,Model Outputs,Errors menu.

When calibrating phenology, coefficients that extend the duration of the life cycle beyond the measured value are often explored. This can result in a model error if the weather data terminate at the measured maturity date. Such a situation can be corrected by duplicating the final days weather and changing the dates for each duplicate at the end of the file. Naturally such additions should be highlighted either by adding some text to the 'Notes' section of the file, or by adding a comment before the added 'artificial' data.

Further, a particular set of calibration coefficients may result in the model running with coefficients not normally encountered. Such 'extreme' coefficient values may 'expose' new model problems, and these should be
communicated to the model developer. To proceed with calibration, the rules file should be changed so that such 'offending' extreme coefficients are not encountered.

1.4 Software Support

GENCALC

L.A.Hunt,
Plant Agriculture,
University of Guelph,
Guelph,
Ontario N1G 1K1,
Canada

thunt@uoguelph.ca (Work)
mthunt@sympatico.ca (Home)

519-824-4120 X53595 (Work)
519-821-2044 (Home)

GenSelect

Paul Wilkens
IFDC
PO Box 2040
Muscle Shoals, AL 35662

256-381-6600
pwilkens@ifdc.org

Index

- G -

- S -

Endnotes 2... (after index)

Back Cover

DSSAT V4.5

Volume 3
Chapter 2

Generalized Likelihood Uncertainty Analysis Tool for Genetic Parameter Estimation (GLUE Tool)

Jianqiang He
Cheryl Porter
Paul Wilkens
Fabio Marin
Howard Hu
James W. Jones

University of Florida, Gainesville, FL
International Center for Soil Fertility and Agricultural Development
International Consortium for Agricultural Systems Applications

Guidelines for Installing and Running GLUE Program

Jianqiang He, Cheryl Porter, Paul Wilkens, Fabio Marin, Howard Hu, and James W. Jones

July 7, 2010

1. Overview

The GLUE (Generalized Likelihood Uncertainty Estimation) program is used to estimate genotype-specific coefficients for the DSSAT crop models. It is a Bayesian estimation method that uses Monte Carlo sampling from prior distributions of the coefficients and a Gaussian likelihood function to determine the best coefficients based on the data that are used in the estimation process. The GLUE program allows users to select a crop, then a cultivar to be estimated. The program will then identify all experiments and treatments in the DSSAT data files for the crop that have measurements for that cultivar. The user then can select one or more experiments and treatments that will actually be used in the coefficient estimation process. Another option for the user is to specify whether to estimate only those coefficients that control phenological development, only those that deal with expansive and dry matter growth, or both sets. Generally, one would want to estimate all parameters. What happens then is that the GLUE program will make 3,000 simulation runs for phenology coefficients and another 3,000 runs for growth coefficients. The program randomly generates parameters that are being estimated (either phenology or growth) from the prior distribution of parameter values and runs the model for each. The model outputs are used to select the parameter set with the maximum likelihood value based on comparison of simulated vs. observed variables, first for phenology parameters, then for growth parameters. The program also computes the uncertainties of the estimates (variances) for each parameter.

The maximum likelihood coefficients are written to a file in the same format as the cultivar file for the selected crop. These values can be copied into the CUL file (e.g., MZCER045.CUL or SBGRO045.CUL, etc.) to operate for routine DSSAT applications and further model evaluations.

What measurements are used to estimate the coefficients? For the development coefficients, measurements of first flower, physiological maturity, and first reproductive organ appearance dates are all used. For growth coefficients, final grain yield, above ground biomass, maximum leaf area during the season, final pod weight, final main stem leaf number, and unit grain weight are used. Thus, the measurements that go into File – A in DSSAT are used; these are variables measured only one time during the season, most of which were measured at harvest.

There are several assumptions that may have important effects on the resulting parameters. First are the prior distributions of coefficients, which are stored in a file called ParameterProperty.xls. This file has information for all of the DSSAT v4.5 crops. We assumed that the parameters have uniform distributions with minimum and

maximum values. This is a conservative assumption, and values are provided in the files based on previous work with the models. A second assumption is that the final errors between simulated and observed values are normally distributed and are unbiased. The assumed values of the variances are given in a file named MeasurementVariances.xls. This assumption may be a problem, particularly if the model is not able to describe responses for a particular experiment very well or if observations are not reliable. Another problem will occur if the experiment had water, nutrient, or other stresses that are either not in the model or that the model does not represent well. Users should only use treatments that are near stress-free conditions, if possible, to minimize these problems. Coefficients estimated using treatments with moderate to severe stress effects will not be reliable. In any case, users should carefully check results from any estimation process to make sure that results are realistic and provide good comparisons to observations used in estimation.

There are other cautions that users should be aware of. For example, results from an estimation process provide conditional estimates of coefficients. That means that the coefficients are the best set given the measurements that were used, but the coefficients also depend on the set of observations used in the process. Our aim is for the coefficients to be robust and useful across environments, but this may not be the case. Another caution is that coefficients estimated from end of season measurements may not reproduce observed time series results very well if such measurements were made. We have seen this occur in various experiments when only end of season measurements are used, whether using GLUE or other estimation procedures. If users have time series data, these data can be used manually to refine the coefficients estimated from the GLUE procedure. It is possible to use in-season measurements and simulations in this type of Bayesian estimation process, but there are certain complications that make it difficult to create a robust and reliable automated procedure.

The GLUE program is one of two tools in DSSAT for estimating cultivar coefficients for the different crops. The first tool, developed by L. A. Hunt and others, evolved from the GENCALC software available in DSSAT v3.5. There are advantages and disadvantages of using each. Disadvantages of the GLUE technique is that it may require a lot of time for the computations, depending on the number of treatments selected for the estimation process. If there are only 2-5 measurement data sets, one would expect the GLUE procedure to finish its calculations in less than 2 hours. However, if there are many measurement data sets, say more than 15, the GLUE method will likely require several hours of calculations. This is a practical limit. On the other hand, the GLUE method can be used, without intervention by users, to produce a set of estimated coefficients. It also provides estimates of the uncertainties of the parameters. This method does not depend on heuristic rules, making it simple to implement for additional crops as they are added to DSSAT.

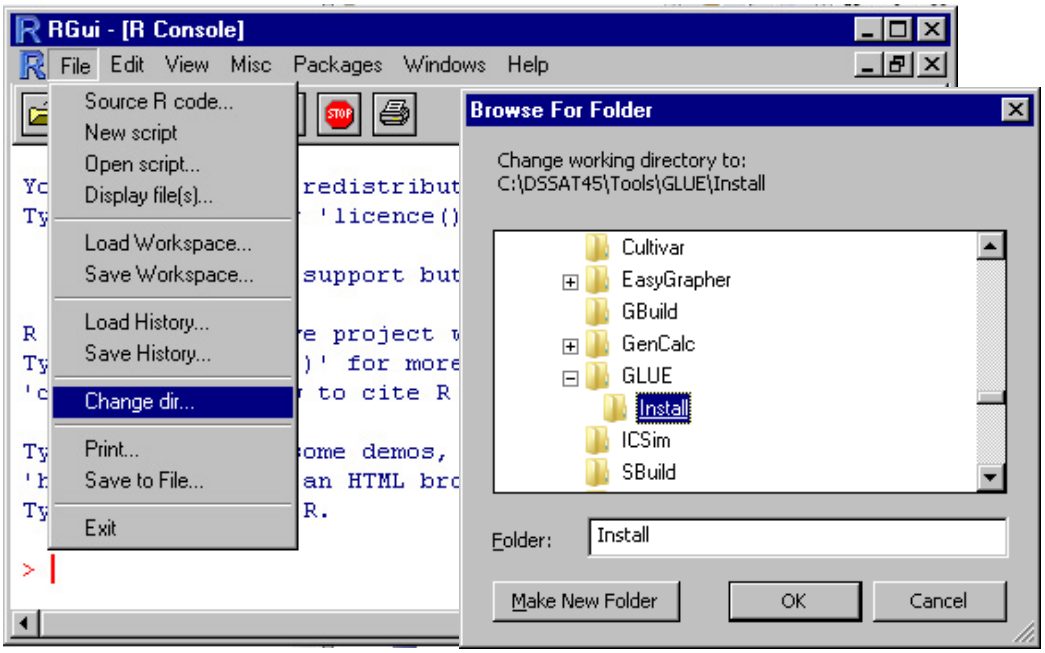
2. Installation of the GLUE Program

The GLUE program was developed using the R statistical programming language. You should have R installed on your computer before proceeding. The DSSAT installation disk has an option to install R, as shown below. You should select the “Install R” button so that R will be available for use in estimating genetic coefficients using the GLUE program.

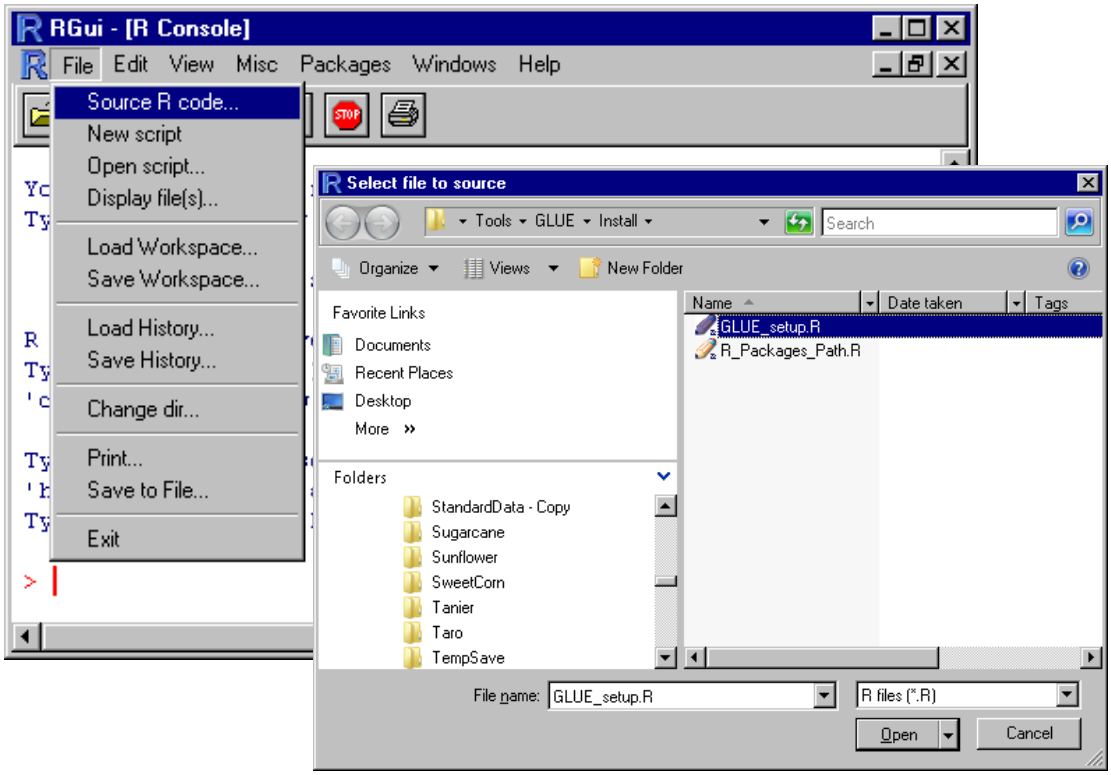
After installing R version 2.10.1 in your machine (from the DSSAT installation disk as shown above), you should configure R to be used with GLUE. The R program requires two packages, “MASS” and “xlsReadWrite”, to run the GLUE program. To install them, follow these steps:

1. Run R with Administrator privileges in sessions where you want to install packages. Do so by right-clicking on the R shortcut and selecting 'Run as Administrator'.

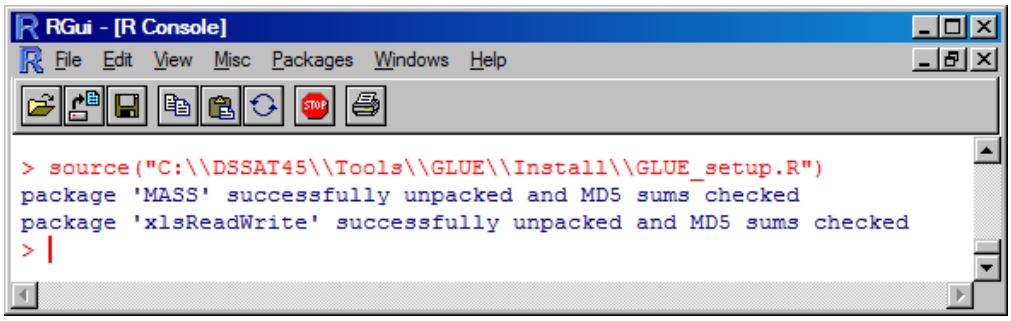
2. Using the File pull-down menu on R, Select “Change dir…” and browse to the DSSAT45/Tools/GLUE/Install directory. Hit “OK” to close the window.



3. Using the File pull-down menu on R, Select “Source R code…” and select GLUE_setup.R and then hit “Open”. By selecting this file, both packages needed for GLUE (“MASS” and “xlsReadWrite”) will be installed.



4. A message will show up in the command space to show you the package has been successfully installed. The package R_packages_path.r does not need to be installed.



5. Close R and return to DSSAT.

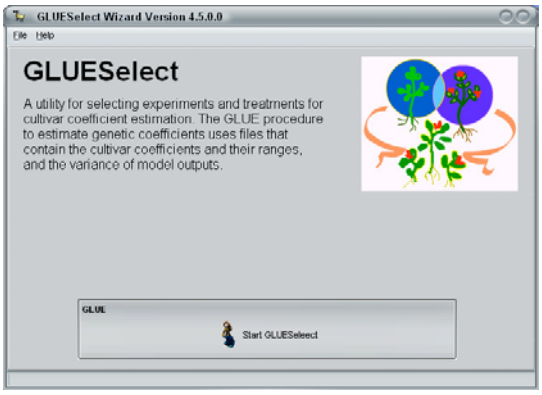
3. Use of the GLUE Procedure to Estimate Genetic Coefficients

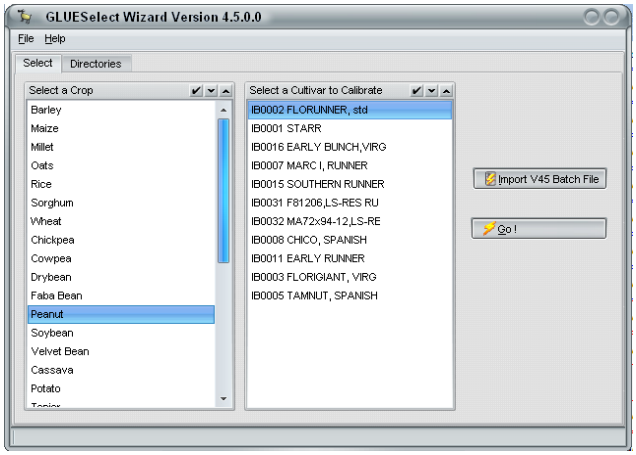
3.1 Setting conditions for GLUE to estimate coefficients

The GlueSelect program was written by Paul Wilkens (IFDC) as a tool in DSSAT v4.5. This tool uses much of the code that he and L. A. Hunt developed for GenSelect, which is a rule-based estimator of cultivar coefficients. Currently, the GLUE program operates on most crops (except those legacy crops that are not converted to v4.5 standards). However, we are more confident of the program correctly estimating cultivar coefficients for the following crops: maize, soybean, peanut, millet, sorghum, chickpea, cotton, fababean, sweet corn, tomato, green beans, rice, wheat, and drybean. Users should check the coefficients carefully before using them. This can be done by putting the estimated parameters in the appropriate CUL file and simulating the crop interactively for comparison with observed data.

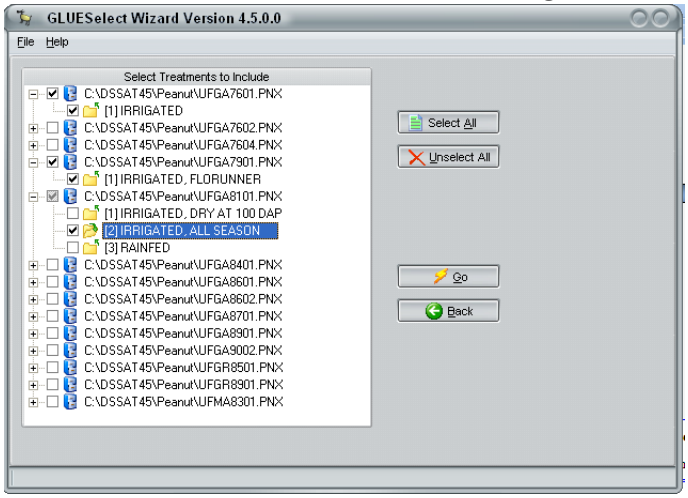
The file that has definitions of genetic coefficients and their ranges of uncertainty for all crops is “ParameterProperty.xls” (see Appendix A for details), and “MeasurementVariance.xls” (see Appendix B). These files are stored in the directory “C:\DSSAT45\Tools\GLUE\”. Advanced model users can modify them to set other ranges of parameters, change parameters to be estimated, introduce parameters for new crops, and change the order in which they are estimated.

The GLUE program is integrated into the DSSAT45 shell, and the user runs the GlueSelect program from the DSSAT Tools menu to start the process as shown at right.

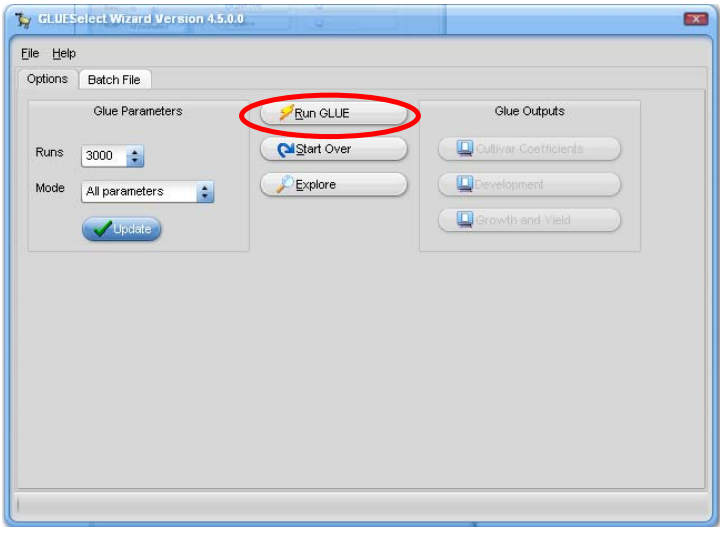




The second GlueSelect screen shows all of the crops. A user selects a crop, such as peanut as shown at left, and then a particular cultivar that is to be estimated ("FLORUNNER, std" in this example). After selecting "Go" on this screen, a list of experiments and treatments will appear as shown below. In this example, three growing seasons were selected from three different years and experiments. These treatments will be simulated, using the Monte Carlo method to estimate the coefficients that give the maximum likelihood for both phenology and growth measurements.



The third GlueSelect screen shows the console for operating the actual GLUE calculations and viewing the results (below). In this example, there are to be 3,000 runs for all parameters, which means that there will be 3,000 runs for phenology parameters and another 3,000 for growth parameters. This number can be changed to a few, say 10, to make sure that the program is operating ok. However, results from runs less than 3,000 would not likely give reliable results. So, if the number is changed from 3,000 to test the procedure, then it must be changed back to 3,000 (or more) in order to get reliable results. It is ok to increase this number to further refine the results, but more time will be required if this is done.



3.2. Running GLUE

The figure above shows how users initiate the simulation runs to estimate the

coefficients using the “Run GLUE” button. Before running GLUE, you may want to disable on-access scanning of files by your anti-virus software. Our experience is that the system will operate much faster when virus scanning is disabled for the DSSAT-generated files. GLUE runs may take some time, possibly from 0.5 to 2 hours for example, depending on how many seasons are selected for estimating the coefficients. From this screen, one can view the final estimated coefficients and copy them to put into the appropriate cultivar file (.CUL in the DSSAT45\Genetics directory) and also to review statistics (mean, maximum likelihood, and standard deviation of the estimated coefficients.

3.3 GLUE Results

The main results that users will be interested in can be seen by selecting the “View Cultivar Coefficient” button on the main screen. This will open an editor with the final values of the estimated coefficients in it. The format of the file is the same as the CUL file for the selected crop, so one can copy this new set of cultivar coefficients into the appropriate CUL file to use in additional simulations. Note that one should use the DSSAT feature to “Update all Lists” after adding a new cultivar to any CUL file.

All of the outputs of DSSAT and GLUE running are saved in the “C:\DSSAT45\GLWork\” directory. The contents of main output files are briefly described as follows:

(a) ***Optimal Parameters.*** The optimal parameter set that was chosen through GLUE procedure was saved as a “CUL” file named according to the name and ID of the selected cultivar when generating the batch file. For example, if the selected cultivar was soybean “COBB”, then the “CUL” file is “SBIB0002 COBB.CUL” (Table 1 in Appendix E).

(b) ***Statistics of Posterior Distributions (Mean, Standard Deviation, and Maximum Likelihood Values).*** The two files identified as “PosteriorDistribution_1.txt” and “PosteriorDistribution_2.txt” (Table 2 and 3 in Appendix E) store the posterior distributions for each round of GLUE, including the mean values, standard deviations, and the parameter set that has the highest likelihood value in that round of GLUE.

(c) ***Empirical Distribution of Parameter Tables.*** The two files identified as “RandomParameterSetsAndProbability_1.txt” and “RandomParameterSetsAnd Probability_2.txt” (Table 4 and 5, Appendix E) store the really used parameter sets and their corresponding probability or normalized likelihood values for each round of GLUE.

(d) ***Generated Parameter Sets.*** The two files identified as “RealRandomSets_1.txt” and “RealRandomSets_2.txt” store the really used parameter set in each round of GLUE.

(e) ***Last Model Run Results.*** “Evaluate_output.txt” stores the content of output file “Evaluate.OUT” of DSSAT for each model run. Since the “Evaluate_output.txt” is processed after each model run, only the result of last model run will be available

in the “Evaluate_output.txt” file after the GLUE procedure. This file is not needed for result analysis, but it is described here because it will be in the directory and model users should ignore it.

(f) ***Results for Computing Likelihood Values.*** The two files “EvaluateFrame_1.txt” and “EvaluateFrame_2.txt” store the appended data of the processed “Evaluate_output.txt” files for the two rounds of GLUE. In each file, the simulated and measured outputs are saved for each treatment and each model run.

(g) ***Combined Likelihood Value for Each Parameter Set.*** The two files identified as “IntegratedLikelihoodMatrix_Frame_1.txt”, and “IntegratedLikelihoodMatrix_Frame_2.txt” (not shown) store the combined likelihood values for all treatments in each model run or for reach parameter set. For example, in the file “IntegratedLikelihoodMatrix_Frame_1.txt”, it stores the combined likelihood values for observations “ADAP”, “MDAP”, and “PD1P” for the first round GLUE. In “IntegratedLikelihoodMatrix_Frame_2.txt”, it stores the combined likelihood values for observations “PWAM”, “HWAM”, “CWAM”, “LAIX”, and “L#SM” for the second round GLUE. When the combined likelihood value is “1” in one column, it means the observation is absent.

(h) ***Combined Likelihood Value for Each Experiment Treatment.*** If there are only two treatments in the experiment for GLUE procedure, then the following files, “IntegratedLikelihoodTreatment_1_1.txt”, “IntegratedLikelihoodTreatment_1_2.txt”, “IntegratedLikelihoodTreatment_2_1.txt”, and “IntegratedLikelihoodTreatment_2_2.txt”, respectively, store the combined likelihood values for each treatment in each round of GLUE. The “IntegratedLikelihoodTreatment_1_1.txt” file, for example, stores the combined likelihood value for GLUE 1 and treatment 1 for all generated parameter sets, so do other files. One can see these files in the DSSAT45/GLWork directory after any GLUE estimation procedure is run.

4. How to Add a New Crop

When a new crop is added to DSSAT, cultivar coefficients for this crop can also be estimated after adding appropriate information in the ParameterProperty.xls file if the naming conventions for measurements and simulated outputs are standardized and the same as for other crops. However, if headers are different for a new crop, then additional information must be added to MeasurementVariances.xls file. The MeasurementVariance.xls file shows how additional sheets in the spreadsheet must be set up for the crop.

Appendices

Appendix A
Example Genotype Parameters and Ranges from ParameterProperty.xls File for Maize and Soybean

	Minimum	Maximum	Flag[1]
MZ	**6**	**6**	**6**
MZ_P1	5	450	1
MZ_P2	0	2	1
MZ_P5	580	999	1
MZ_G2	248	990	2
MZ_G3	5	16.5	2
MZ_PHINT	49	49	0
SB	**18**	**18**	**18**
SB_CSDL	11.78	14.6	0
SB_PPSEN	0.100	0.385	1
SB_EM.FL	9	23.5	1
SB_FL.SH	10	10	0
SB_FL.SD	12	16	1
SB_SD.PM	29	37.7	1
SB_FL.LF	18	18	0
SB_LFMAX	0.95	1.15	2
SB_SLAVR	300	400	2
SB_SIZLF	140	230	2
SB_XFRT	1	1	0
SB_WTPSD	0.158	0.195	2
SB_SFDUR	17	25.5	2
SB_SDPDV	1.7	2.44	2
SB_PODUR	10	10	0
SB_THRSH	78	78	0
SB_SDPRO	0.4	0.4	0
SB_SDLIP	0.2	0.2	0

[1]The FLAG column indicates which coefficients are to be estimated using phenology measurements (FLAG=1), which are to be estimated using growth measurements (FLAG=2) and which coefficients are not to be estimated (FLAG=0).

Appendix B
Variances of Observations for Most Crops

	STD	Variance	CV	Flag	Description
ADAP	3	9		1	Anthesis day (dap).
MDAP	7	49		1	Physiological maturity day (dap).
PD1T	4	16		1	First pod date (YrDoy).
PWAM			0.3	2	Pod/Ear/Panicle weight at maturity (kg [dm]/ha).
HWAM			0.3	2	Yield at harvest maturity (kg [dm]/ha).
CWAM			0.3	2	Tops weight at maturity (kg [dm]/ha).
LAIX			0.4	2	Leaf area index, maximum.
HWUM			0.1	2	Grain unit weight at maturity (g/seed)
L.SM	3	9		2	Leaf number per stem at maturity. The symbol "#" was changed to ".", since it is the symbol of comments in R.

Appendix C
Batch File “COBB.SBC” Created with GLUESelect

```
$BATCH(CULTIVAR):SBIB0002 COBB

@FILEX                                          TRTNO       RP
SQ        OP        CO
C:\DSSAT45\Soybean\UFGA8101.SBX                     1        0
0         0         0
C:\DSSAT45\Soybean\UFGA8501.SBX                     1        0
0         0         0
```

Appendix D
Batch File “GLUE.BAT”

```
C:\Progra~1\R\R-2.10.1\bin\Rterm                --slave                 <
C:\DSSAT45\Tools\GLUE\GLUE.r
```

Appendix E
Output Files of GLUE Procedure

1. Optimal parameter set saved as a CUL file (SBIB0002 COBB.CUL)

```
IB0002 COBB (8)          . SB0801 12.54 0.373 23.83 9.200 20.11 32.63 18.00 1.090 346.0 162.7 1.000 0.184 17.16 1.846 10.00 78.00 0.400 0.200
```

2. Posterior distribution in first round GLUE (PosteriorDistribution_1.csv)

Param	SB_CSDL	SB_PPSEN	SB_EM.FL	SB_FL.SH	SB_FL.SD	SB_SD.PM	SB_FL.LF	SB_LFMAX	SB_SLAVR	SB_SIZLF	SB_XFRT	SB_WTPSD	SB_SFDUR	SB_SDPDV	SB_PODUR	SB_THRSH	SB_SDPRO	SB_SDLIP
Mean	12.174	0.303	23.286	9.2	18.558	31.511	18	1.03	375	190	1	0.158	23	1.9	10	78	0.4	0.2
STDEV	0.327	0.073	3.602	0	2.587	2.99	0	0	0	0	0	0	0	0	0	0	0	0
MaxProbability	12.538	0.373	23.831	9.2	20.113	32.635	18	1.03	375	190	1	0.158	23	1.9	10	78	0.4	0.2

3. Posterior distribution in second round GLUE (PosteriorDistribution_2.csv)

Param	SB_CSDL	SB_PPSEN	SB_EM.FL	SB_FL.SH	SB_FL.SD	SB_SD.PM	SB_FL.LF	SB_LFMAX	SB_SLAVR	SB_SIZLF	SB_XFRT	SB_WTPSD	SB_SFDUR	SB_SDPDV	SB_PODUR	SB_THRSH	SB_SDPRO	SB_SDLIP
Mean	12.538	0.373	23.831	9.2	20.113	32.635	18	1.187	345.327	178.795	1	0.184	18.651	2.042	10	78	0.4	0.2
STDEV	0	0	0	0	0	0	0	0.105	25.85	27.183	0	0.008	1.324	0.22	0	0	0	0
MaxProbability	12.538	0.373	23.831	9.2	20.113	32.635	18	1.09	346.041	162.69	1	0.184	17.165	1.846	10	78	0.4	0.2

4. Example random parameter sets and their Likelihood values and in first round GLUE (RandomParameterSetsAndProbability_1.txt)

CSDL	PPSEN	EM.FL	FL.SH	FL.SD	SD.PM	FL.LF	LFMAX	SLAVR	SIZLF	XFRT	WTPSD	SFDUR	SDPDV	PODUR	THRSH	SDPRO	SDLIP	Probability
12.538	0.373	23.831	9.200	20.113	32.635	18.000	1.030	375.000	190.000	1.000	0.158	23.000	1.900	10.000	78.000	0.400	0.200	0.227
12.549	0.364	26.064	9.200	21.883	29.549	18.000	1.030	375.000	190.000	1.000	0.158	23.000	1.900	10.000	78.000	0.400	0.200	0.117
11.791	0.200	26.731	9.200	15.802	34.909	18.000	1.030	375.000	190.000	1.000	0.158	23.000	1.900	10.000	78.000	0.400	0.200	0.115
11.895	0.187	27.931	9.200	16.092	34.792	18.000	1.030	375.000	190.000	1.000	0.158	23.000	1.900	10.000	78.000	0.400	0.200	0.103
11.784	0.262	21.206	9.200	13.915	29.413	18.000	1.030	375.000	190.000	1.000	0.158	23.000	1.900	10.000	78.000	0.400	0.200	0.095
12.339	0.376	19.642	9.200	19.395	28.400	18.000	1.030	375.000	190.000	1.000	0.158	23.000	1.900	10.000	78.000	0.400	0.200	0.085

5. Example random parameter sets and their Likelihood values and in second round GLUE (RandomParameterSetsAndProbability_2.txt)

CSDL	PPSEN	EM.FL	FL.SH	FL.SD	SD.PM	FL.LF	LFMAX	SLAVR	SIZLF	XFRT	WTPSD	SFDUR	SDPDV	PODUR	THRSH	SDPRO	SDLIP	Probability
12.538	0.373	23.831	9.200	20.113	32.635	18.000	1.090	346.041	162.690	1.000	0.184	17.165	1.846	10.000	78.000	0.400	0.200	0.016
12.538	0.373	23.831	9.200	20.113	32.635	18.000	1.029	342.632	189.759	1.000	0.186	17.602	1.775	10.000	78.000	0.400	0.200	0.015
12.538	0.373	23.831	9.200	20.113	32.635	18.000	1.110	330.573	174.342	1.000	0.186	17.317	2.364	10.000	78.000	0.400	0.200	0.015
12.538	0.373	23.831	9.200	20.113	32.635	18.000	1.011	307.951	203.991	1.000	0.190	17.585	2.005	10.000	78.000	0.400	0.200	0.015
12.538	0.373	23.831	9.200	20.113	32.635	18.000	1.082	339.885	225.893	1.000	0.194	18.123	1.792	10.000	78.000	0.400	0.200	0.015
12.538	0.373	23.831	9.200	20.113	32.635	18.000	1.150	375.298	163.908	1.000	0.191	17.539	1.719	10.000	78.000	0.400	0.200	0.014

Volume 3
Chapter 3

Introductory Crop Simulation Tool (ICSIM)

O. Uryasev
J. W. Jones
Mamoud Medany
Gerrit Hoogenboom
Suat Irmak
Cheryl Porter

University of Florida, Gainesville, FL
Central Laboratory for Agricultural Climate, Cairo, Egypt
The University of Georgia, Griffin, GA
Washington State University, Prosser, WA
International Consortium for Agricultural Systems Applications

ICSim

Introductory Crop Simulation Tool for DSSAT v4[1]

Oxana Uryasev[2]
Research Associate

James W. Jones[2]
Distinguished Professor

Mahmoud Medany[3]
Head, Agrometeorological Application Dept.

Gerrit Hoogenboom[4]
Professor

Suat Irmak[2]
Graduate Research Associate

Cheryl Porter[2]
Coordinator, Computer Applications

Research Report ABE 2000-04
Agricultural and Biological Engineering Department
University of Florida
Gainesville, FL 32611

April 1, 2003

[1] This software was developed as part of the project entitled "Integrated Crop Management Information System for Crops, [ICMIS], A Decision Information System for Crops [DISC]" under the ATUT Collaborative Research funded by the USDA/FAS/ICD Research and Scientific Exchange Division.

[2] Agricultural and Biological Engineering Department, University of Florida, Gainesville, FL 32611.
[3] Central Laboratory for Agricultural Climate, Agricultural Research Center, Ministry of Agriculture, Cairo, Egypt.
[4] Department of Biological and Agricultural Engineering, The University of Georgia, Griffin, Georgia 30223-1797

Table of Contents

1 Purpose

The purpose of **ICSim** is to introduce new users to some of the main capabilities of crop simulation models of DSSATv4 software package. In particular, it allows users to specify weather, soil, and main practices to simulate crop growth and yield using any of the crop models and data sets contained in DSSAT v4. The program also automatically computes potential yield, assuming no limitations in water and nitrogen for the same inputs. This provides new users an immediate demonstration of the concept of yield gap analysis by comparing potential yields with yields that may be limited by water or nitrogen or both.

2 General Description

ICSim is a MS Windows application program that is used as an add-on to DSSAT v4 (ICSim is compatible with DSSAT v4). ICSim makes use of the DSSAT directory structure and communication protocol, which designates the location of all programs and data files used in DSSAT. The Visual Basic (VB version 6) program was developed as part of a US-Egypt cooperative project and was conceived after introducing DSSAT to students in Egypt during a one-week training program. We learned that students were sometimes intimidated and confused by the numerous options and capabilities in DSSAT, which seemed to be a deterrent to understanding its basic concepts and uses. Thus, this experience demonstrated the need for windows-based software that would introduce first-time users to the basic ideas and capabilities of crop models first, then allow them to move into its advanced capabilities more slowly.

Note: *A more comprehensive windows-based tool has also been developed that will provide access to all of the functionality of the crop models. The crop simulation tool for DSSAT4 (XBuild) allows users to specify any combination of management options for simulation of several crops for purposes of validation (comparison with observed data), seasonal analysis, crop rotations, and spatial analysis that are available in DSSAT v4.*

ICSim uses the DSSAT v4 files, directories, and crop models, and therefore the documentation available in Tsuji et al. (1994) and Hoogenbom at al. (1999) (volume 4) applies to this software. The only major difference between this software and DSSAT 4 is that ICSim creates its own working directory (called \DSSAT4\ICSim) for all management files and for all outputs of crop models. This allows beginning users to learn about the operation of DSSAT without endangering the existing files and other programs.

The purpose of this report is to provide a description of the installation of ICSim and a tutorial demonstrating its use.

3 System Requirements

Hardware:

PC with at least 35 MB of free disk space

Software:

Windows 98/NT/2000/XP Operating System
DSSAT v4

4 Installation

1. Before installation, close all applications on your computer.
2. Insert the installation CD and run the setup program. Windows will execute the setup program and displays two preliminary dialog boxes. Then the program will display the dialog box in Fig.1. Although the setup program lets the user select the directory to install the program, it is recommended that the default directory be used. The default directory for installing ICSim is C:\DSSAT4\Tools\ICSim\. After clicking on the button shown in Figure 1, the installation begins.

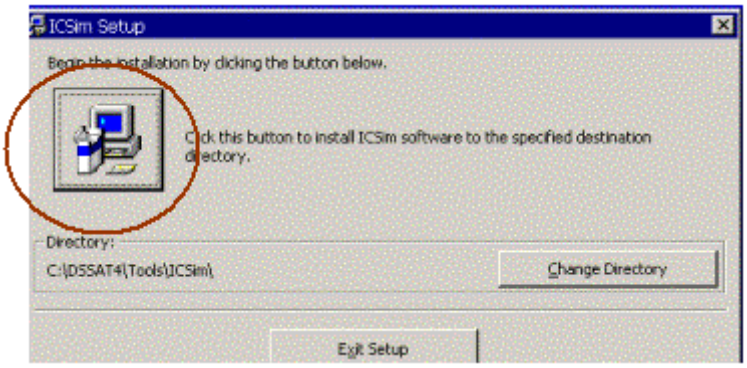


Figure 1. Window displayed by Setup.exe, prompting the user to click the button to begin installation, and offering an opportunity to change installation directory.

Important: ICSim must be installed on the same drive as DSSAT4!

During the installation, the program registers all OCXs (ActiveX controls) and copies all necessary files. If some of the files already exist on your system, warning messages will be displayed. It is recommended that you keep the newest version of the files when you are prompted.

After the installation is completed, ICSim is ready to use without rebooting the computer.

5 Tutorial User's Guide

5.1 Getting Started

Short-cut: You can create an icon for ICSim and bring it to the desktop. To do this, highlight the ICSim.exe application file (in the directory where it was installed), then "Right click" on the mouse and choose the **Create Shortcut** option from the menu. Rename this shortcut as **ICSim**, cut it, and paste it to your desktop.

Menu Items: ICSim is much like many other Windows applications. It has a menu bar at the

top of its main window. With the following main menu options:

- **File**
- **Environmental Modifications**
- **Management**
- **Simulation**
- **Help**

The **File** menu option allows user to
(1) Create a **new** scenario for simulation;
(2) **Open** an existing file for simulating a scenario previously defined;
(3) Display a simulation scenario while it is being edited (**print preview**);
(4) **Print**;
(5) **Save** the current scenario.

The simulation scenarios define the weather location and year, the soil, and management practices for simulation. These scenarios are stored in "experiment" files(X-files), using the naming convention of DSSAT v3.5 (Tsuji *et al.*, 1998). All files created by ICSim are automatically stored in the working directory named \DSSAT4\ICSim.

The **Environmental Modification** menu option allows the user to make changes to daily weather variables that are read from a weather file. The user can thus run simulations by modifying environmental variables such as daily total radiation, maximum temperature, minimum temperature, precipitation, CO_2.

The **Management** menu option allows the user to enter management inputs for planting, irrigation applications, fertilizer applications, and harvesting dates.

The **Simulation** menu option runs the appropriate crop model in DSSAT v4, then allows users to display and view output results and to print and graph model outputs.

5.2 New Simulation

To define a completely new simulation scenario, go to **New** under the **File** menu option. The **Experimental Details** window will be presented as shown in Figure 2. When the screen first appears, most of the slots are empty. Default soil initial conditions for percent available water and nitrogen are shown, but users can change these. This screen is called **Experimental Details**.

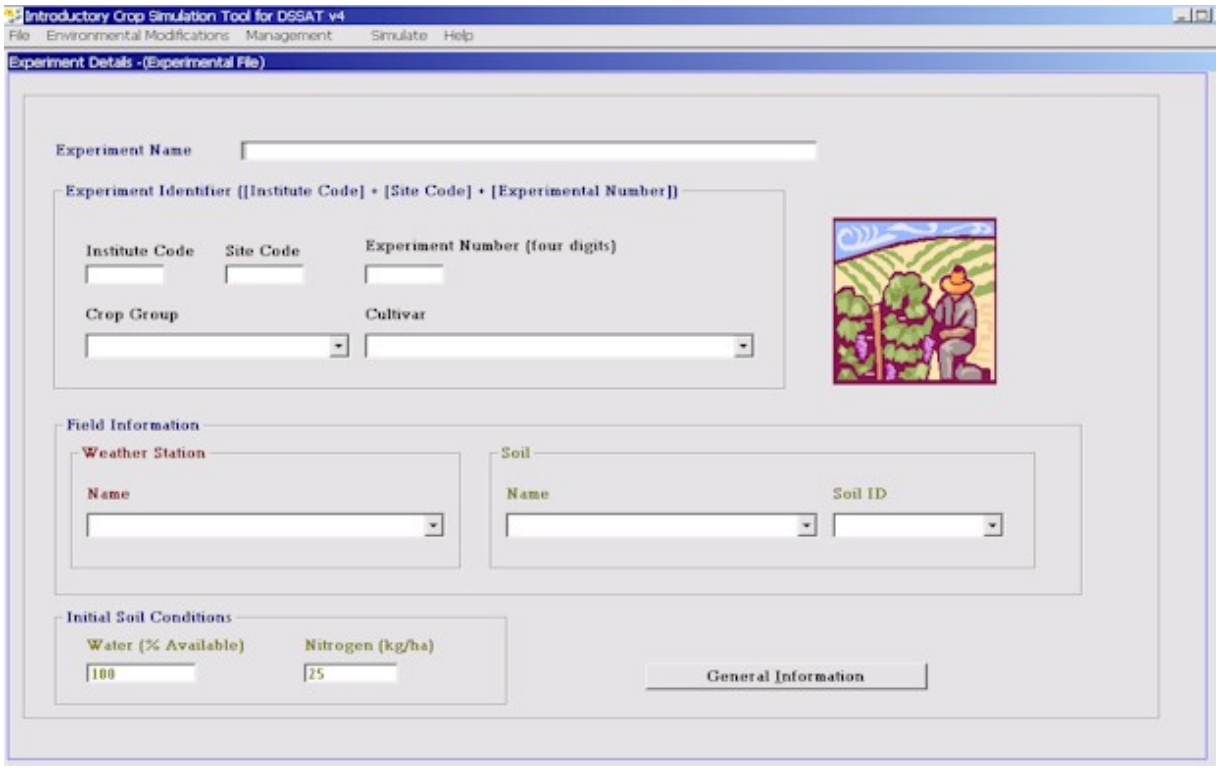


Figure 2. This is the first screen that allows users to define a new simulation scenario.

5.2.1 Experimental Details

The **Experimental Details** window contains **Experiment Name**, **Experiment Identifier**, **Field Information**, and **Initial Soil Conditions** options (Fig. 2).

The user can enter a name for the scenario (or experiment) in the **Experiment Name** box. Up to 60 characters can be entered to identify the experiment to be simulated.

The **Experiment Identifier** is made up of a 2-character **Institute code**, a 2-character **Site code,** and a 4-character **Experiment code**. The experiment identifier is used to name the file that will contain the scenario that you are creating. This identifier will also become a header in the file to identify the experiment code and name. The experiment code uses the same convention as the DSSAT4 file naming system; it provides information on institute, site, planting year, experiment number, and crop. For example, UFGA9901.MZX, is the name of the experimental (X) file for maize (MZ) experiment 01, planted in 1999 by institute "University of Florida" (UF) at site "Gainesville" (GA). This information must be entered into the corresponding boxes on the screen.

The crop to be simulated is selected by a pull-down menu on this screen. After selecting the crop, one must select a cultivar from its pull down menu.

Note: *The program creates/modifies only* ***experimental*** *(FILEX) files.*

The scenario (and thus the experiment file) can also contain the names of the people supplying the data set and information on the plot sizes, etc., used in the experiment (**General Information** button). These latter items are not normally used by simulations models, but are provided for reference and assistance in interpreting simulation results.

Field Information (Weather Station and **Soil):** Pull-down menus list all available observed **weather** data. The user selects the weather name (Figure 2). If the user wants to simulate crop growth using different weather data than are presented in this menu, a new weather file has to be added to \DSSAT4\Weather directory. Please check the DSSAT documentation for detailed information on file content, format, and naming convention.

A pull-down menu lists all available **Soil** data. The user selects the **Soil Name**.

Initial Soil Conditions (% Water and Nitrogen): This option allows the user to enter the percent available water content on the day of planting. The default value is 100%, which means that all soil layers have **water** content at the drained upper limit value. Users may change this (between 0 and 100%); the value of 0 sets all soil layers to the lower soil **Water** limit at the start of simulation. Also the user will enter the initial **Nitrogen** (kg/ha) in the soil profile on the day of planting. The default value is 25 kg/ha, which is distributed evenly over all soil depths. Users can change this (the value has to be greater than 0.0).

Tip: *Initial values of* ***water*** *and* ***nitrogen*** *are assumed to be uniform throughout the soil profile; users cannot modify the individual values of each soil layer.*

Once inputs are completed in **Experimental Details** window, the user will go to next step to define crop management.

5.2.2 Management Information

The **Management** window is displayed to enter management data for planting date, plant population, row spacing, irrigation, fertilizer and harvesting date inputs.

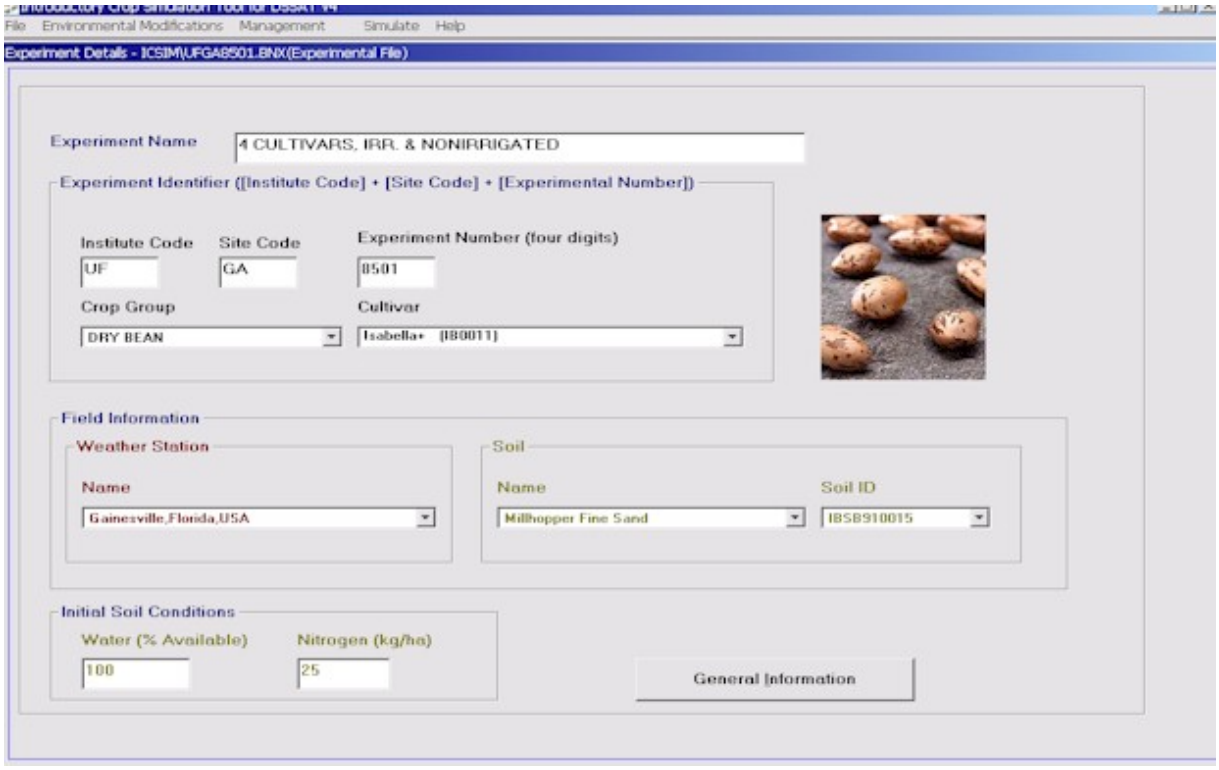


Figure 3. Management input screen for specifying planting date, plant population, row spacing, irrigation, fertilizer and harvesting date inputs.

Planting: User-supplied planting information that users must enter includes planting date, plant population and row spacing, as shown in Figure 3. Planting date must be in the order of month, day of year and last two-digit of year using the following format: (MM/DD/YY or MM/DD/YEAR). The planning year must be selected from menu **Available Weather Years** by clicking on the corresponding year in the list.

Note: *This planting* ***date*** *will be used to set the dates for initial condition measurement, start simulation date, and environmental modification date (if this section is used).*

Tip: *The planting year can be entered to the corresponding box manually. You must enter only a year, which is listed in the* ***Available Weather Years*** *list, otherwise an error message will be issued. When entering manually, you may enter two last digits for a year. When the year entered as two digits, is less then 10, it is assumed that it is in the 21st century. For example, the date 03/04/10 is the same as 03/04/2010 and date 03/04/65 is 03/04/1965.*

Irrigation Application: If the user wishes to include irrigation, then **irrigation application date**, **irrigation type** and **amount** must be entered in the irrigation application frame (Figure 3). Each row in the irrigation table corresponds to a single irrigation application. The user must enter the irrigation application date to the corresponding cell in correct format.

Irrigation amount can be entered by clicking the mouse into the **Amount, depth of water, mm** cell in the third column. A value of "0" is the default value.

To select different irrigation **type**, click on the corresponding cell in the irrigation application table and choose the desired irrigation type from the pull-down list.

To add more **irrigation applications** for your simulations, click on the **Add** button at the top right of the irrigation table. A new row will be added at the bottom of existing row. To select the default values for this new entry, the user must first select an existing row in the table, and then click the **Add** button. Then users can change inputs for the new irrigation application by changing the values in the row.

Tip: *User can add a new irrigation application in the middle of two existing irrigation applications (i.e. between 12th and 13th irrigation applications). To do this, "Highlight" the first irrigation option (12th row), and click on the* ***Add*** *button. A new row will be added after 12th row. User can edit the date, irrigation amount and type for the new row (it is 13th row now!). User can use the scroll bar to go up and down in the* ***Irrigation Application*** *frame for editing. This scroll bar will appear, if there are more than 5 irrigation applications. If the desired application is not "Highlighted", ICSim will add a new irrigation just after the last irrigation (last row).*

To **delete** an existing irrigation application, highlight the desired entry that you want to delete. Then, click on the **Delete** button on the right side of the irrigation table. A warning message box will pop up by asking "Are you sure you want to delete an application of chosen date?" Click the **Yes** button, and the selected irrigation application will be removed.

Note: *ICSim sorts the irrigation applications by date of application each time when the* ***Management*** *window is loaded. Users should apply one irrigation for a day. However, when simulating paddy (flooded) rice, users will need to enter more than one row to specify water level, height of the field dike (or bund) surrounding the field, and percolation rate on the first day of the crop. Users are referred to the DSSAT V3.5 User's Guide, Volume 2 for a description of the extra inputs required for this crop.*

Fertilizer Application: Users can add **fertilizer application date**, **fertilizer type** and **amount** in the **Fertilizer Application** table. Each fertilizer application event requires one row in this table.
Fertilizer **date** in the first column must be in the format of MM/DD/YY (or MM/DD/YEAR). Users can click the mouse in the **Date** cell and enter the date of a fertilizer application. **Fertilizer amount** can be entered by clicking the mouse into the **N applied fertilizer, kg/ha** box.

To select different **Fertilizer type**, double click the **Type** cell in the **Fertilizer** table, and choose the desired fertilizer **type** from the pull-down menu (i.e. Ammonium Sulfate).

To enter additional fertilizer application events, click on the **Add** button at the top right of the fertilizer table. To select the default values for this new entry, the user must first select a row in the table, and then click the **Add** button. Users can change the inputs for new fertilizer application by changing the values in the row.

To delete an existing fertilizer application, highlight the desired entry that you want to delete.

Then click the **Delete** button on the right of the fertilizer table. A warning message box will pop up by asking "Are you sure you want to delete an application of chosen date?". Click **Yes** to remove the selected fertilizer application.

Note: *ICSim sorts the fertilizer entries by date of application each time when the* ***Management*** *window is loaded. Users should apply only one fertilizer on any particular day.*

Harvesting: This section defines the final harvest date. Setting the final harvest date as the maturity date can do this or the user can define a harvest date in the format of MM/DD/YY (or MM/DD/YEAR) in Date cell. To enter a date or to remove it, the user must choose the option At Maturity or At Specific Date. Harvest date is required for some crops to run the simulations.

5.2.3 Environmental Modifications

When users click on the **Environmental Modification** menu item, the **Environmental Modification** window will be displayed (Figure 4). This window allows users to change weather conditions for the crop simulation. A default starting date for modifying daily weather is already entered in the **Date** cell at the top of the window. The format to enter the date is MM/DD/YY (or MM/DD/YEAR). Any day after this date, daily weather values will be changed according to the modifications entered into the other slots on this window. The user can change this date to be sometime during the season. This allows one to simulate an environment switching experiment, such as a shade experiment imposed 40 days after planting.

The **Environmental Modification** window allows users to modify up to 8 different **environmental** variables for simulation. These environmental variables and their units are: **Daylength** (hour); **Radiation** (MJ m^{-2} day^{-1}); **Maximum daily temperature** (^{o}C); **Minimum daily temperature** (^{o}C); **Precipitation** (mm); (6) **CO_2** (vpm); (7) **Humidity** (%); and (8) **Wind speed** (km d^{-1}).

Tip: *The default date for* ***environmental modifications*** *is the planting date. Thus, it is advisable to enter management information, including planting date, before opening the "Environmental Modification" window.*

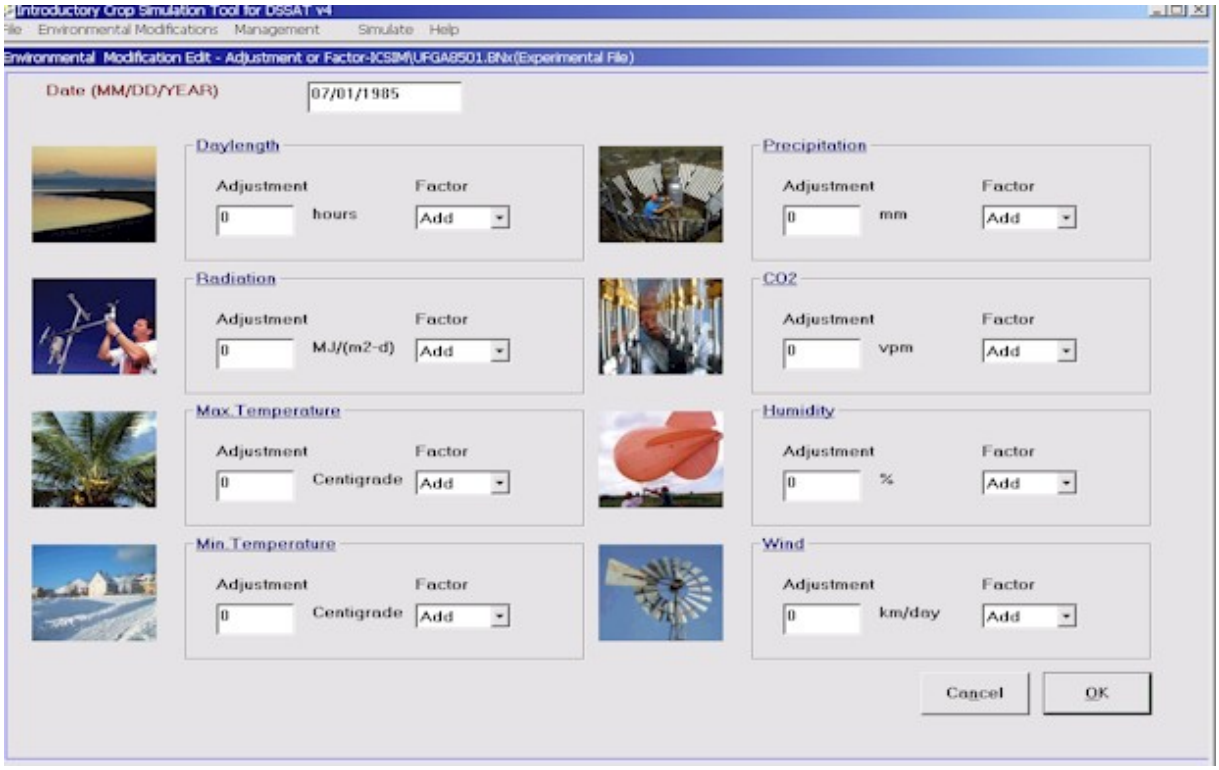


Figure 4. Environmental Modification Screen.

The name of each environmental variable is shown at the top of its frame. Each environmental variable (i.e. Radiation) includes an **Adjustment** box and a **Factor** pull-down menu. The **Adjustment** box allows users to enter a value for adjusting daily values of that variable (i.e. Radiation). The **Factor** pull-down menu provides 3 menu options: "Add", "Multiply" and "Replace". For an environmental variable (i.e. **Radiation**), users can enter a value in **Adjustment** box and specify the type of adjustment using menu items in the **Factor** pull-down menu. For instance, to simulate results under a 3 C^o hotter temperature regime starting on day MM/DD/YY (or MM/DD/YEAR), the user would enter the date in **Date** box, and then enter a value of 3 in the **Adjustment** boxes of both **Max. Temperature** and **Min. Temperature** cells, and finally select item "Add" from the **Factor** pull-down menu

Note: *Users can only specify one environmental modification in ICSim. This modification is assumed to start on the date specified, which is defaulted to the planting date. Once activated, the* ***environmental modification*** *remains in effect throughout the season.*

5.2.4 Print Preview

The print preview option is available under the **File** menu in ICSim. When users click on the **Print Preview** option, the **File Preview** window will appear (Fig. 5). Print preview allows users to display an experiment file while they are being created or edited. Using the scroll bar on the right corner in **File Preview** window (Fig. 5), users can go to the bottom of the file. Users have to click on the "Refresh" button to see current changes. To close print preview, click on the **OK** button.

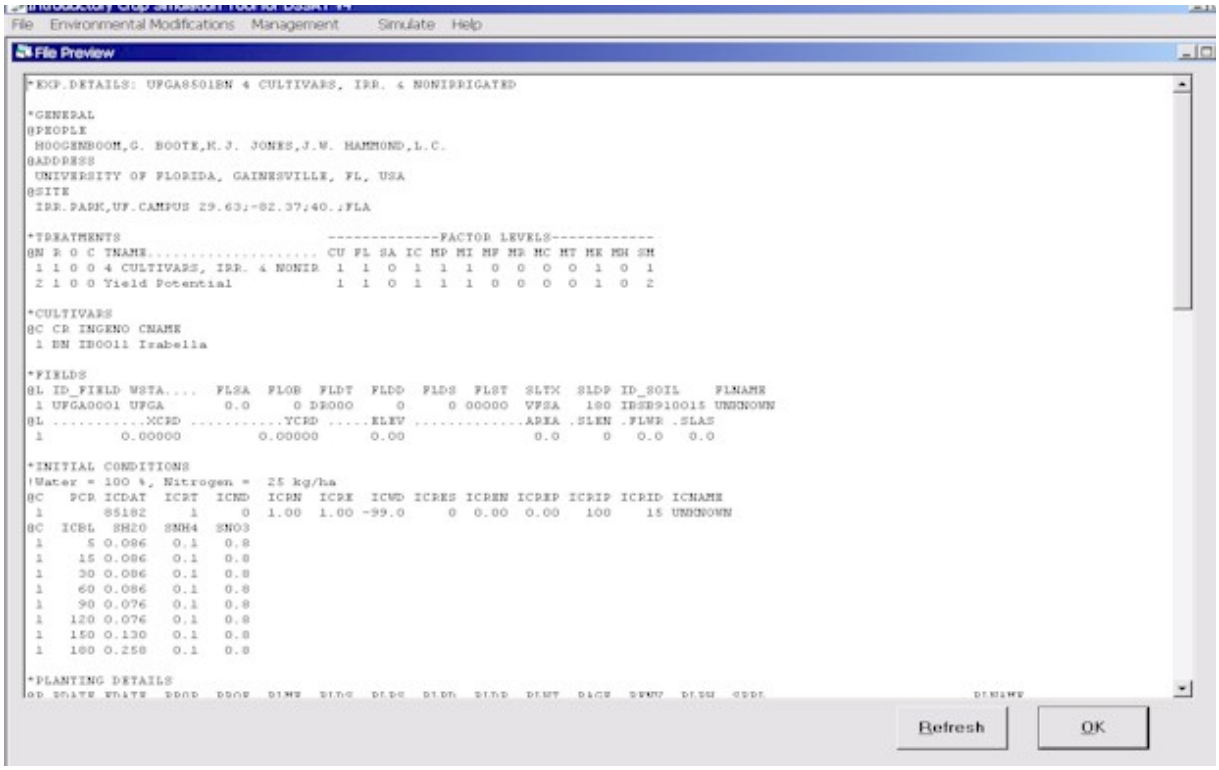


Figure 5. Display of the file being created or edited, using the print preview selection under **File** in the main toolbar of ICSim.

5.2.5 Running the Simulation

By clicking on the **Simulation** menu button, the file being created or edited will be saved and the crop selected will be **simulated** using the crop models in DSSAT v4.

Tip: *The simulation outputs will be overwritten once a new simulation is done. User has to rename the output files using a file manager, if output files intended to be used later.*

Note: *In ICSim, users do not have full access to all simulation options in DSSAT v4. For example, users cannot define automatic management characteristics for planting, irrigation, residue, and harvest. ICSim uses default simulation controls (from a file named DEFAULT_FILE.TXT), which is copied to your computer during installation into the directory where ICSIM is installed. You can open DEFAULT_FILE.TXT using any text editor to change defaults. However, do not modify this or other files without first making a backup copy.*

5.2.6 Analyzing Results

When the simulation is finished, a window with options for viewing, printing, and graphing results will appear (Fig. 6). There are different buttons for analyzing results in the **Simulated Results** window (i.e. Graph, Print etc.).

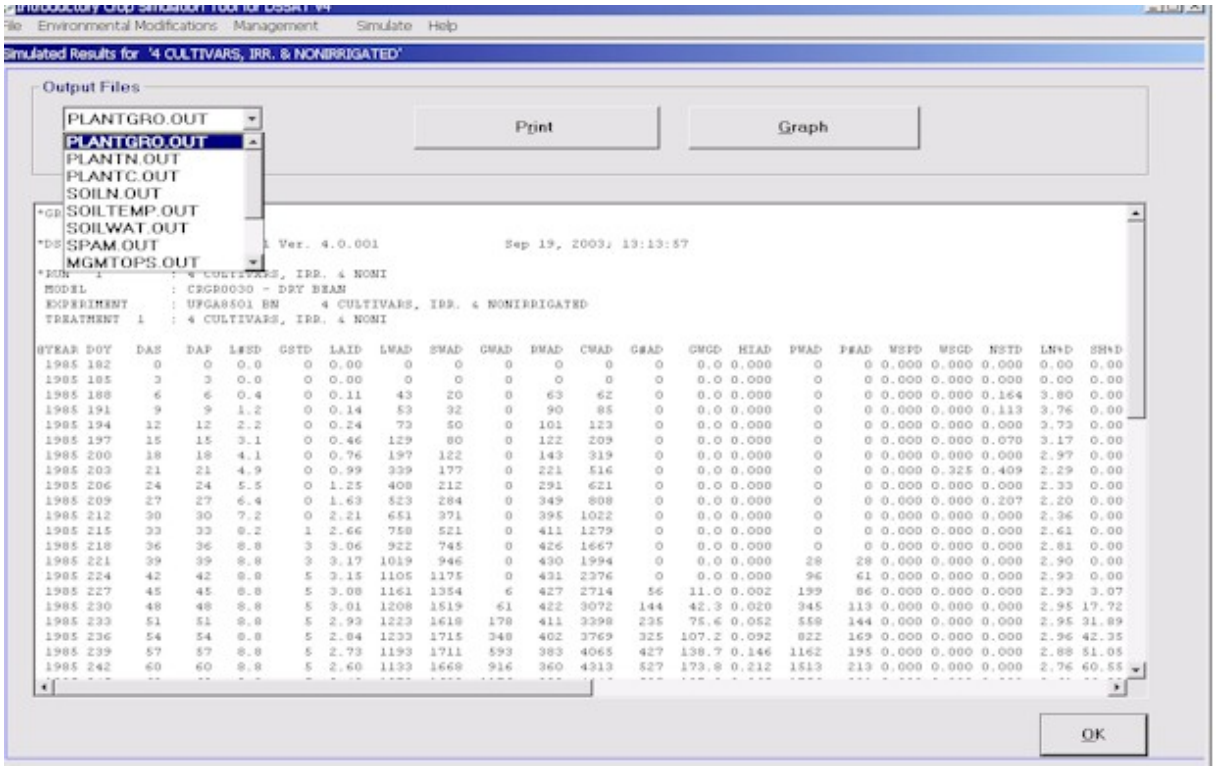


Figure 6. Simulated results screen that provides users with the capability to view, print, and graph results.

Viewing Model Outputs: Once each simulation is run, a number of output files are written. Pull down menu lists these files. The selected file is displayed in the textbox bellow. The user can use the scroll bar to view any part of the selected output file. A short description of each output file is given in "Inputs and Outputs" section in the tutorial.

Printing Model Outputs: The **Print File** button allows users to print the output file that is currently displayed in **Simulated Results** window.

Graphing Model Outputs using DSSAT v4 Graphics Package: All output files, except OVERVIEW.OUT and SUMMARY.OUT can be plotted. The Graph button activates DSSAT4 graphic utility (GBuild). This allows users to display outputs as a function of day of year or planting date. After the graph program is opened, the user selects the variables to plot (Fig. 7).

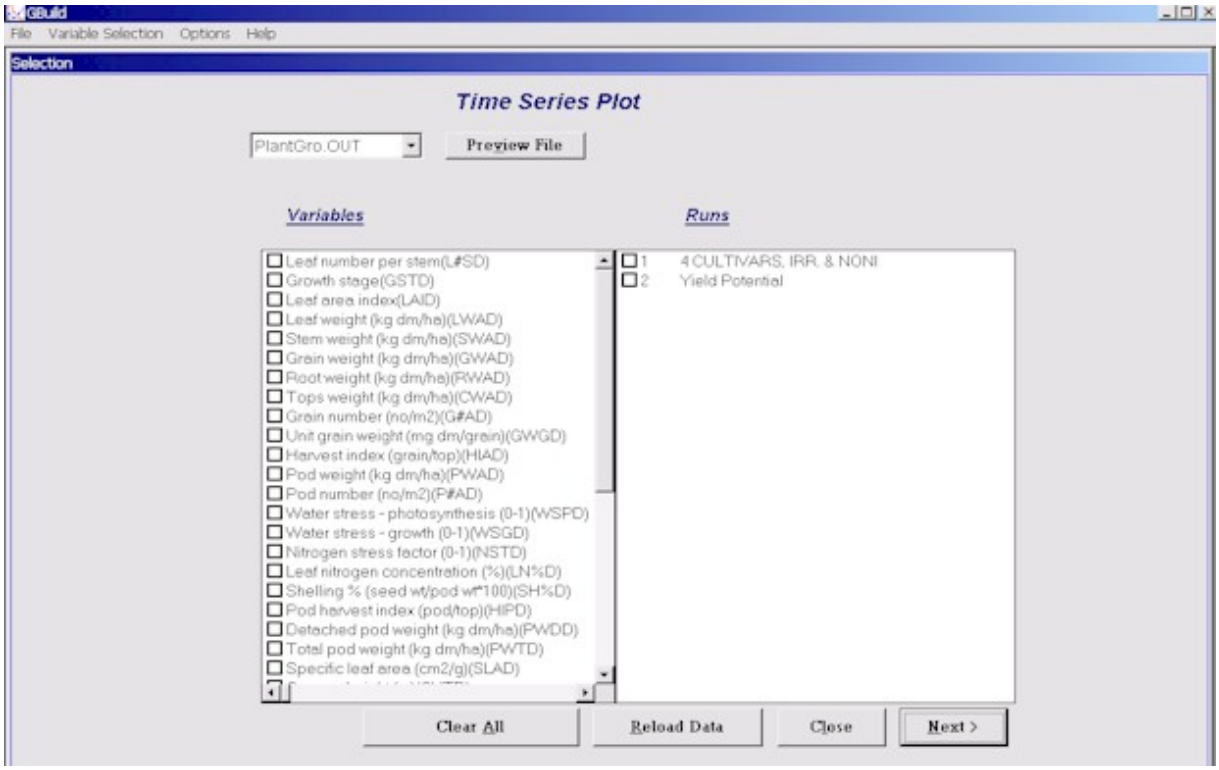


Figure 7. Screen shot of GBuild.exe showing some of the simulated variables from the selected output file (here PLANGRO.OUT) that can be graphed.

Note: *In ICSim, graphs are produced for simulated results only. The graphs do not contain experimentally measured values of variables. Omission of the measured results was intentional for simplification and introductory purposes.*

The button **Next** will open the show graph for the selected variables.

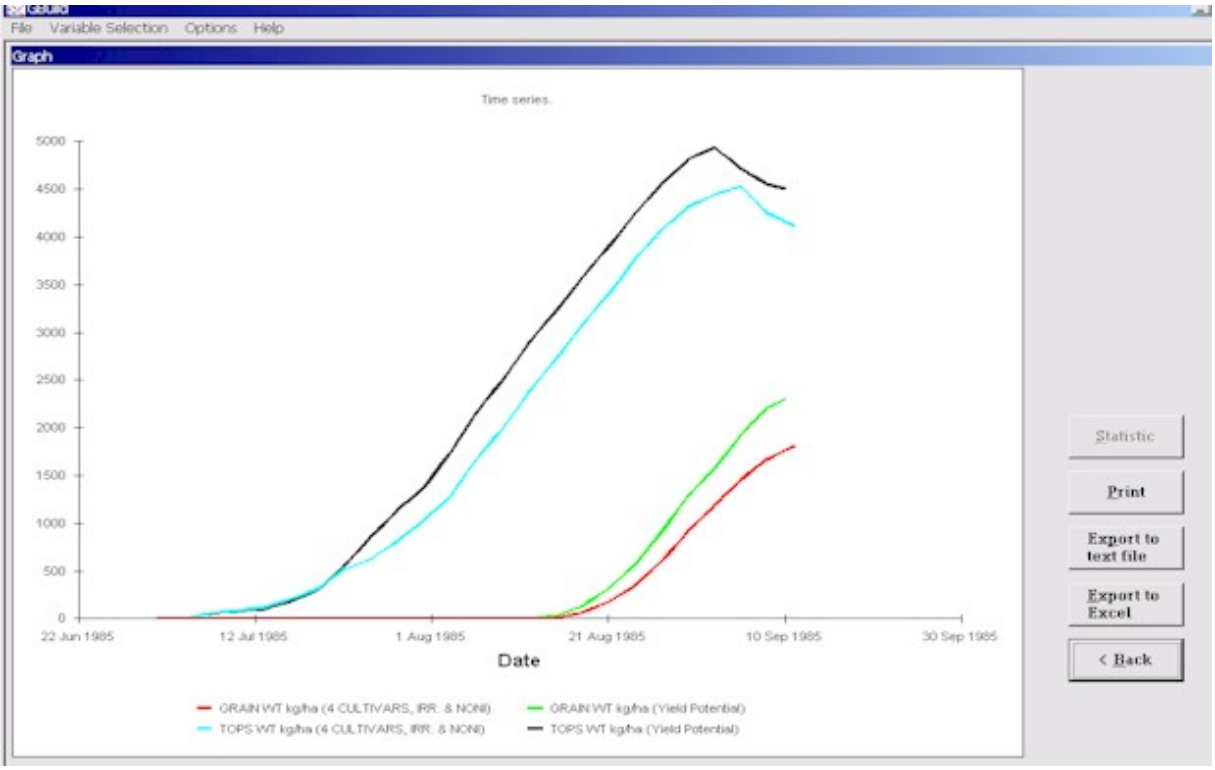


Figure 8. Graphic results for the selected variables.

5.3 Opening an Existing File to Setup a Simulation

To setup a simulation run, ICSim provides capabilities for users to open an existing file. Under **File** menu, click on **Open**.

Note: *By default, the initial directory in the open files dialog box is ICSim. If this folder does not exist, the initial directory will be DSSAT 4.*
Files located in other than ICSim folder can be used as examples for experimental files. The program ***does not*** *modify any of these files. Although the user can make any changes on the screen, ones the file is saved, it will be stored in ICSim folder. If this folder does not exist in the system, it will be created automatically.*

Once the experiment file is loaded, users can edit the information using **Experimental Details**, **Environmental Modification** and **Management** windows. Clicking **OK** in the **Experimental Details** window will save the file in the \DSSAT4\ICSim directory. ICSim saves the experiment file with the name that is defined in **Experimental Details** window. When a user clicks on the **Simulation** menu, the model will run using the inputs shown in the different windows and will allow one to display outputs from this run just as for a new file.

5.4 Files

The ICSim input/output file structures are the same as those in DSSAT v4, which is described in Volume 2-1 (Jones et al., 1994) of the DSSAT user's manual. Short descriptions of these

files are given below.

5.4.1 Inputs

Weather: The ICSim uses the weather files located in the path DSSAT4\WEATHER. In DSSAT, the name of the weather file is constructed internally from the weather file information contained in the experimental input file. The code for the weather file includes eight characters (four letters + four numbers), such as UFGA6001, and the file UFGA6001.WTH would be constructed in the model for a simulation run. The daily weather data file includes site name for the station, number of years of the data for this site, latitude, longitude, and elevation. The daily weather input data includes solar radiation (SRAD), maximum and minimum temperatures (TMAX and TMIN), rain (RAIN).). Part of the weather file UFGA6001.WTH is shown in Figure 9.

Soil: ICSim uses the **soil surface** and **soil profile** information located in DSSAT4\Soil\SOIL.SOL file. The soil file includes soil depth and soil properties for each layer of a soil profile. The main soil properties are: **drained upper** and **lower limits**, **runoff curve number**, **drainage, soil texture** (% sand and clay), **bulk density, pH, organic carbon, hydraulic conductivity**, etc. (for more details see, Volume 2-1 (Jones et al., 1994) of the DSSAT user's manual). Figure 10 shows characteristics for a soil from the SOIL.SOL file in DSSAT v4.

```
*WEATHER DATA : Gainesville,Florida,USA

@ INSI      LAT     LONG  ELEV   TAV   AMP REFHT WNDHT
  UFGA   29.630  -82.370    10  20.9   7.5  2.00  3.00
@DATE  SRAD  TMAX  TMIN  RAIN                 PAR
60001   6.5  16.7   8.9  13.2                13.2
60002   9.4  24.4  13.9   0.0                19.1
60003   7.1  23.3  18.3  13.5                14.4
60004  13.1  20.6   5.6   0.0                26.6
60005  12.8  26.1  11.1   0.0                26.1
60006   9.9  27.2  13.9   0.0                20.0
60007   7.2  23.9  16.1   2.0                14.5
60008  13.1  19.4   4.4   0.0                26.7
60009  12.5  23.3   3.9   0.0                25.4
60010  12.8  25.6   7.8   0.0                26.0
60011  13.5  24.4   9.4   0.0                27.3
60012  11.5  25.0   6.7   0.0                23.3
60013  13.2  26.7   9.4   0.0                26.7
60014   8.4  26.7  12.8   0.0                17.1
60015  13.6  25.6  13.9   0.0                27.6
60016  14.1  25.6   5.6   0.0                28.7
60017  13.5  24.4   3.9   0.0                27.4
60018  11.8  21.1  13.9   3.8                23.9
60019  14.4  17.2   2.8   0.0                29.2
60020  13.9  10.0  -2.2   0.0                28.3
60021  14.6   8.9  -3.9   0.0                29.7
60022  14.7  12.2  -5.6   0.0                29.9
60023  14.9  12.2  -5.0   0.0                30.1
```

Figure 9. Example weather file that is used in ICSim, showing the first 24 days of the daily weather data for Gainesville, Florida for 1960 from the file named UFGA6001.WTH.

```
*IB00000005  IBSNAT      SIL0    150 DEFAULT - MEDIUM SILTY LOAM
@SITE        COUNTRY          LAT     LONG SCS FAMILY
 Generic     Generic          -99      -99 Generic
@ SCOM  SALB  SLU1  SLDR  SLRO  SLNF  SLPF  SMHB  SMPX  SMKE
   -99  0.12   6.0  0.30  79.0  1.00  1.00 IB001 IB001 IB001
@  SLB  SLMH  SLLL  SDUL  SSAT  SRGF  SSKS  SBDM  SLOC  SLCL  SLSI  SLCF  SLNI  SLHW  SLHB
     5   -99  .106  .262  .462 1.000   -99  1.37  1.16    10    60     0  .116   6.5   -99
    15   -99  .106  .262  .462 1.000   -99  1.37  1.16    10    60     0  .116   6.5   -99
    30   -99  .106  .262  .462 0.819   -99  1.37  1.10    10    60     0  .110   6.5   -99
    45   -99  .107  .262  .462 0.607   -99  1.37  0.97    10    60     0  .097   6.5   -99
    60   -99  .107  .262  .462 0.607   -99  1.37  0.97    10    60     0  .097   6.5   -99
    90   -99  .108  .261  .461 0.368   -99  1.38  0.72    10    60     0  .072   6.5   -99
   120   -99  .110  .260  .460 0.202   -99  1.38  0.43    10    60     0  .043   6.5   -99
   150   -99  .111  .259  .459 0.111   -99  1.39  0.20    10    60     0  .020   6.5   -99
```

Figure 10. Example soil parameters required by the crop models.

5.4.2 Outputs

During each simulation run, several output files are created. Their format is the same as those in DSSAT v4. Outputs are written to the following files:

OVERVIEW.OUT: This output file provides an overview of input conditions and simulated crop performance. Results of ICSim simulation runs are not linked with experimental measurements for validation. Therefore, the observed values that are in the OVERVIEW.OUT file will all be missing (missing data are shown by the value of –99). The first section in this file presents information that uniquely describes the simulated data set. This for instance includes Experiment name, treatment number, crop, planting date, population, and row spacing, soil initial conditions, etc.

The second section in OVERVIEW.OUT contains a summary of soil characteristics and cultivar coefficients. The next section deals with crop and soil status at the main development stages, followed by a comparison of simulated and measured data for major variables. This in turn, is followed by information on simulated stress factors and weather data summary during the different development phases (as appropriate to the crop).

SUMMARY.OUT: This file provides a summary of outputs for use in applications programs with one line of data for each crop season. It shortly contains a one-line summary of the main development events, water and nitrogen variables, and yield components. It has two lines of information; one for the scenario defined by the user simulated with water and nitrogen limitations considered and the other for potential yield, simulated with the water and nitrogen balances turned off.

6 Differences and Limitations Relative to DSSAT v4 Capabilities

There are inherent differences between the operation of this introductory tool (ICSim) for DSSAT v4 and that of the full (XBuild) DSSAT v4. These differences were necessary to provide new users an introduction to the basic features of the crop models in DSSAT v4 without having to immediately learn all of the features available in DSSAT v4.

1. ICSim provides capabilities to set up simulation runs for any of the crop models in DSSAT4 for single season simulations only. It does not allow users to simulate a number of years to estimate the effects of annual weather variability on crop production.
2. Results of ICSim simulation runs are not linked with experimental measurements for validation. Therefore, the observed values that are usually printed in the OVERVIEW.OUT file are omited.
3. ICSim simulates the crop, soil and management conditions specified by the user using the crop models operating in the water and nitrogen limitation mode (i.e., with the soil water and nitrogen balances turned on). It also automatically simulates the potential yield (ie. With no water or nitrogen stresses) for the same scenario so that users can easily gain an impression of the yield gap associated with that weather, soil, and management situation.
4. If a user opens an existing file for any crop, only the first treatment in that file is used. The original file that is opened is NOT modified. Instead, a new file is written to the ICSim operating directory (DSSAT4\ICSim).This new file remains in the ICSim directory until the user deletes it.
5. In ICSim, soil, water and nitrogen initial conditions are set in a simple way on the first screen. Users specify the available soil water (in percentage) and initial soil nitrogen (kg/ha). These values are assumed to exist on the date of planting specified by the user in the **Management** screen. In addition, these values are assumed to be uniform throughout the soil profile. In ICSim, users do not have the opportunity to modify the individual values for all of the soil layers.
6. Initial residue for simulations done with ICSim is assumed to be 1000 kg/ha, and cannot be changed by the user.
7. In the **Environmental Modification** screen, users can only specify one modification. This modification is assumed to start on the date specified in the corresponded box on the screen, which is defaulted to the planting date. Once activated, the environmental modification remains in effect throughout the season.
8. ICSim does not allow for pest damage to be imposed on the crops.
9. ICSim does not provide users with access to seasonal, sequence, or spatial analysis programs in DSSAT v4, nor does it provide access to other important utilities, such as WeatherMan (Hansen et al., 1994) or GENCALC (Hunt et al., 1993).

7 References

Chan, H., P. W. Wilkens, G. Hoogenboom, J. W. Jones, and D. C. Godwin. 1994.Graphing Simulated and Experiment Data. In DSSAT version 3. A Decision Support System for Agrotechnology Transfer. Volume 2. G. Y. Tsuji, G. Uehara, and S. Balas (eds.). University of Hawaii. Honolulu, HI. pp. 245-281.

Hansen, J. W., N. Pickering, J. W. Jones, C. Wells, H. Chan, and D. C. Godwin. 1994. WeatherMan. In DSSAT version 3. A Decision Support System for Agrotechnology Transfer. Volume 3. G. Y. Tsuji, G. Uehara, and S. Balas (eds.). University of Hawaii. Honolulu, HI. pp. 137-199.

Hoogenboom, G., J. W. Jones, P. W. Wilkens, W. D. Batchelor, W. T. Bowen, L. A. Hunt, N. B. Pickering, U. Singh, D. C. Godwin, B. Baer, K. J. Boote, J. T. Ritchie, and J. W. White. 1994. Crop Models. In DSSAT version 3. A Decision Support System for Agrotechnology Transfer. Volume 2. G. Y. Tsuji, G. Uehara, and S. Balas (eds.). University of Hawaii. Honolulu, HI. pp. 95-244.

Hunt, L. A., S. Pararyasingham, J. W. Jones, G. Hoogenboom, D. T. Imamura, and R. M. Ogoshi. 1993. GENCALC - Software to Facilitate the use of Crop Models for Analyzing Field Experiments. Agron J. 85:1090-1094.

Hunt, L. A., J. W. Jones, P. K. Thornton, G. Hoogenboom, D. T. Imamura, G. Y. Tsuji, and U. Singh. 1994. Accessing Data, Models and Application Programs. In DSSAT version 3. A Decision Support System for Agrotechnology Transfer. Volume 1. G. Y. Tsuji, G. Uehara, and S. Balas (eds.). University of Hawaii. Honolulu, HI. pp. 21-110.

Imamura, D. T., J. W. Jones, P. K. Thornton, and G. Hoogenboom. 1994. Creating Management Files to run Models and Document Experiments. In DSSAT version 3. A Decision Support System for Agrotechnology Transfer. Volume 1. G. Y. Tsuji, G. Uehara, and S. Balas (eds.). University of Hawaii. Honolulu, HI. pp. 111-164.

Jones, J. W., L. A. Hunt, G. Hoogenboom, D. C. Godwin, U. Singh, G. Y. Tsuji, N. B. Pickering, P. K. Thornton, W. T. Bowen, K. J. Boote, and J. T. Ritchie. 1994. Input and Output Files. In DSSAT version 3. A Decision Support System for Agrotechnology Transfer. Volume 2. G. Y. Tsuji, G. Uehara, and S. Balas (eds.). University of Hawaii. Honolulu, HI. pp. 1-94.

Tsuji, G. Y., J. W. Jones, G. Hoogenboom, L. A. Hunt, and P. K. Thornton. 1994. Introduction. In DSSAT version 3. A Decision Support System for Agrotechnology Transfer. Volume 1. G. Y. Tsuji, G. Uehara, and (eds.). University of Hawaii. Honolulu, HI. pp. 1-11.

Tsuji, G. Y., G. Uehara, S. Balas. 1999. DSSAT version 3. Volume 1-3

G. Hoogenboom, P. W. Wilkens, G. Y. Tsuji. 1999. DSSAT version 3. Volume 4

DSSAT V4.5

Volume 3
Chapter 4

Statistical Evaluation (Stats)
and
Weather Analogue Program (Ana)

A. S. du Toit
D. L. du Toit

University of Florida, Gainesville, FL
Sustainable Farming Systems, Potchefstroom, South Africa
International Consortium for Agricultural Systems Applications

Short Description of the Model Statistical Package and Weather Analogue Program

A.S. Du Toit[1] & D.L. du Toit[2]

[1] ARC Grain Crops Institute, Potchefstroom, RSA
[2] Sustainable Farming System, Kock Park, RSA

Abstract

To facilitate model validation and adaptation in the Highveld Ecoregion Project, two software packages were developed. The Model Statistical Package (MSP) uses standard outputs of CERES-Maize 3.0 to calculate indicators of model accuracy, including the linear regression statistics (slope, intercept, and r^2), D-index, and the systematic and unsystematic mean square errors. The Weather Analogue Program (WAP) allows users to create mid-season projections of crop performance based on the five sets of historical weather that show the greatest similarity to the ongoing season.

Introduction

The Highveld ecoregion of southern Africa is an important producer of rainfed maize but its highly variable rainfall patterns present a major challenge to producers. Simulation models offer various possibilities for guiding farmers on managing production risk. Two objectives of the Highveld Ecoregion Project[1] are "to validate existing ICASA models and adapt them to be applicable for agricultural systems in the Highveld Ecoregion." To facilitate model testing and application of models in the Highveld region, two software tools were developed to analyze outputs from the DSSAT group of models. These two tools were the Model Statistical Package (MSP) and the Weather Analogue Program (WAP). This paper describes these two tools, using CERES-Maize v3.0 as part of DSSAT models (Hoogenboom et al. 1994) to illustrate the software features. Such tools would also be of great use in evaluating proposed improvements in the temperature response of crop models.

The CERES Model

CERES models for maize, sorghum, wheat, millet, and barley have been combined, resulting in a generic multi-crop CERES3. This version runs from a single executable fileset of code, incorporating the development and growth sections for each individual model into a single module with a single soil component (Tsuji et al. 1994). According to Ritchie (1991), generic models should allow users to follow more uniform procedures for validating models and for linking them with components not included in the generic model. The generic CERES3 (Hoogenboom et al. 1994) was used for the simulations, with modifications as reported by Du Toit (1996).

Model Statistical Package

In order to verify and calibrate a model, well-defined criteria are needed to evaluate model performance. It is generally accepted that the ultimate test of a simulation model is the accuracy with which it describes or mimics the actual system, usually involving comparisons between simulated and observed data (Willmott 1982; Jones and Kiniry 1986; Oreskes et al. 1994). A number of statistical methods for analysing model performance are available. These include linear regressions techniques as proposed by Jones and Kiniry (1986) and Flavella (1992) and D-index, systematic and unsystematic mean square errors as proposed by Willmott (1982).

[1] Funded by the Ecoregion Fund Managed by ISNAR

Jones and Kiniry (1986) used linear regression techniques of the form $y = a + bx$ with simulated results as the independent variable. Good model performance was obtained when the intercept (b) approached zero and the slope of the regression approached unity, indicating a near perfect relationship between observed and simulated values. Complementary to this regression, the Pearson correlation coefficient (R) can also be calculated, indicating the similarity or inverse similarity of a response in y for a response in x. The coefficient of determination (r^2) is readily calculated from R, signifying the percentage of variation that is accounted for by the model.

The deficiencies of the latter statistical parameters were pointed out by Willmott (1982) and Harrison (1990). They indicated that the observed and simulated data might occur in a narrow band, whereas this is usually not the case with the coefficients. Secondly, although significance levels can be calculated, it is difficult to identify the point when a model is valid or not valid. Savage (1993) further warned against the use of a correlation coefficient if the data are not randomly distributed.

The D-index (index of agreement), RMSEs (root mean square error systematic), RMSEu (root mean square error unsystematic) and RMSE (root mean square error) are four indicators that Willmott (1982) recommended for model evaluation. Due to limitations in the use of correlation coefficients as an agreement index, Savage (1993) stated that the statistics as defined by Willmott (1982) should be used instead.

According to Wilmott (1982), the D-index of a "good" model should approach unity and the RMSEs approach zero, whereas the RMSEu should approach the RMSE. The mean absolute error (MAE) and RMSE are among the best overall measures of model performance, as they summarise the mean difference in the units of observed and predicted (Willmott 1982):

$$MAE = N^{-1} \sum_{i=1}^{N} |Pi - Oi| \qquad \text{Eq. 1}$$

$$RMSE = \left[N^{-1} \sum_{i=1}^{n} (Pi-Oi)^2 \right]^{0.5} \qquad \text{Eq. 2}$$

N is the number of observed values and Oi and Pi are observed and predicted values for the i-th data pair. The systemic and unsystemic RMSE require calculating the intercept (a) and slope (b) of the least-squares regression, $\hat{P} = a + b0\ Pi$ (Willmott 1982),

$$RMSEs = \left[N^{-1} \sum_{i=1}^{n} \left(\hat{P}i - Oi \right)^2 \right]^{0.5} \qquad \text{Eq. 3}$$

$$RMSEu = \left[N^{-1} \sum_{i=1}^{n} \left(Pi - \hat{P}i \right)^2 \right]^{0.5} \qquad \text{Eq. 4}$$

where $\hat{P}i$ is regarded as the best estimate of the predicted quantity (Savage 1993). The advantage of RMSEs is that it indicates the bias (deviation of the actual slope value from the 1:1 line) in a particular model, instead of the random variation (RMSEu) that may occur (Savage 1993). Willmott (1982) proposed an "index of agreement" (D) of the form:

$$D = 1 - \left[\sum_{i=1}^{n} (Pi-Oi)^2 / \sum_{i=1}^{n} \left(|P'i| + |O'i| \right)^2 \right] \qquad \text{Eq. 5}$$

where $P'i = Pi - O$ (average of the observed) and $O'i = Oi - O$. The index (D) is intended to be a descriptive measure. It is also both a relative and bounded measure, which can be widely applied in order to make cross-comparisons between models (Willmott 1982)

Observed vs. simulated graphs, also known as 1:1 graphs, are widely used in simulation evaluations (Willmott 1982; Jones and Kiniry 1986). Harrison (1990) recommended that it is appropriate to combine whatever statistical method is used with a 1:1 graph since it may be particularly helpful in identifying the pattern of differences between the predicted and observed values.

The MSP software reports all of the above-mentioned statistical parameters using the "`MAIN GROWTH AND DEVELOPMENT VARIABLES`" provided in the file `Overview.out` by comparing the predicted with the measured values as indicated in Table 1. The measured values are then subtracted from the predicted values; the difference is the error or residual. A correlation matrix is calculated between the error and each of the 50 predicted environmental and stress factors as indicated in Table 2. The correlation matrix is presented as a regression matrix table by MSP as indicated in Figure 1.

Table 1. Main growth and development variables in the Overview.out file as output from a model run with CERES3.

@	VARIABLE	PREDICTED	MEASURED
	FLOWERING DATE (dap)	81	-99
	PHYSIOL. MATURITY (dap)	160	-99
	GRAIN YIELD (kg/ha)	**5882**	**5853**
	WT. PER GRAIN (g)	0.4016	
	GRAIN NUMBER (GRAIN/m2)	1237	
	GRAINS/EAR	824.99	
	MAXIMUM LAI (m2/m2)	1.20	
	BIOMASS (kg/ha) AT ANTHESIS	2441	-99
	BIOMASS N (kg N/ha) AT ANTHESI	0	-99
	BIOMASS (kg/ha) AT HARVEST MAT.	8244	

Table 2. ENVIRONMENTAL AND STRESS FACTORS in the `Overview.out` file as output from a single run with CERES3.

ENVIRONMENT						STRESS			
DEVELOPMENT PHASE	TIME		WEATHER			WATER		NITROGEN	
	DURATION days	TEMP MAX oC	TEMP MIN oC	SOLAR RAD MJ/m2	PHOTOP [day] hr	PHOTO SYNTH	LEAF EXPAN.	PHOTO SYNTH	LEAF EXPAN.
Emergence-End Juvenile	15	28.27	11.67	20.86	13.47	0.000	0.000	0.000	0.000
End Juvenil-Floral Init	7	25.90	12.90	22.23	13.64	0.000	0.000	0.000	0.000
Floral Init-End Lf Grow	49	22.62	13.10	22.68	13.72	0.000	0.000	0.000	0.000
End Lf Grth-Beg Grn Fil	16	25.45	12.31	20.63	13.37	0.000	0.000	0.000	0.000
Grain Filling Phase	62	23.05	13.60	21.18	12.38	0.000	0.000	0.000	0.000

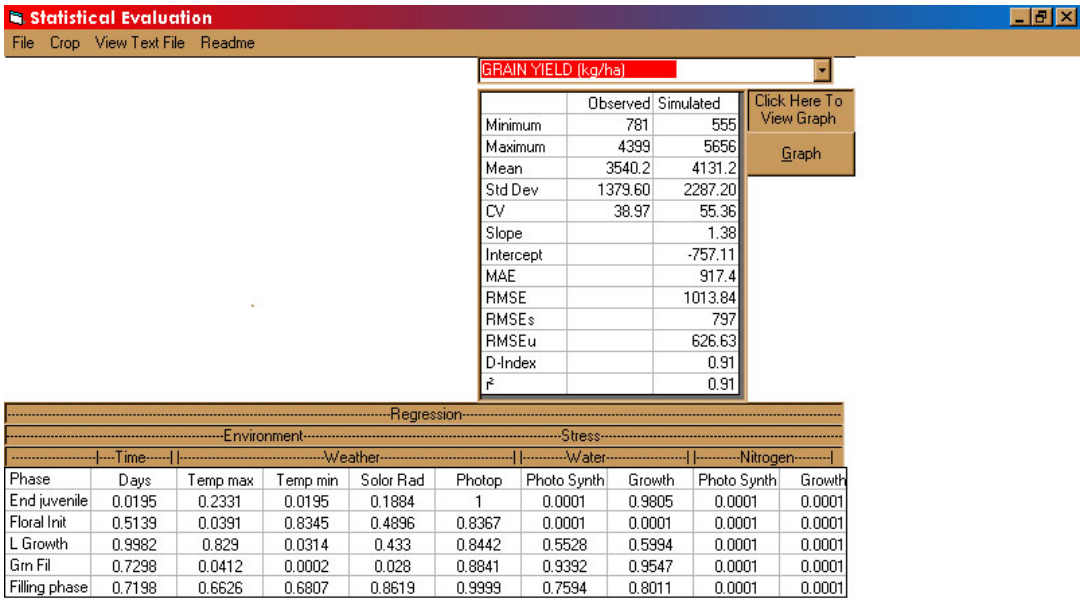


Figure 1. Output of the MSP to indicate the statistical parameters and relationship between the environmental and stress factors and the difference between the observed and the simulated.

Weather Analogue Program

The Weather Analogue Program *(WAP)* was developed to assist seasonal predictions modelled for a given site. WAP uses an analogue methodology whereby the most current weather conditions of the present season are compared to the five best fitting historical seasons in the database. The information is then presented in the form of a graph and a table. The search algorithm uses the "index of agreement" (D) as represented in Eq. 5 (Willmott 1982) but compares daily weather data instead of observed or simulated crop performance. Each season is presented as the simulation output of a model with the current season as the observed values. Figure 2 shows the graph indicating the two closest years. Although they are presented as cumulative values, the search algorithm uses non-cumulative data.

The data are indicated as cumulative on the graph to help with interpretation. WAP finds the five models that best fit the weather data of the current season as indicated in Figure 3.

This program has a user-friendly spreadsheet interface for data input and stores data in standard DSSAT v3 and 3.5 ASCII format. The most recent data of the current season are appended to the historical weather datasets as indicated in Figure 4.

The program also exports five files, consisting of the most recent weather data of the current season combined with daily data for the five analogue years that are used to complete the season. The program can be run via the keyboard or from an ASCII file used within a GIS framework.

CERES runs from a control file that describes the experimental simulation trial. For South Africa, such a trial consisted of six plantings at two-week intervals in combination with super short, short, medium and long growing season cultivars. All of these were in combination with a low medium and high plant population. Each of these treatments was then run for all five of the analogue weather files exported by WAP (Figure 5).

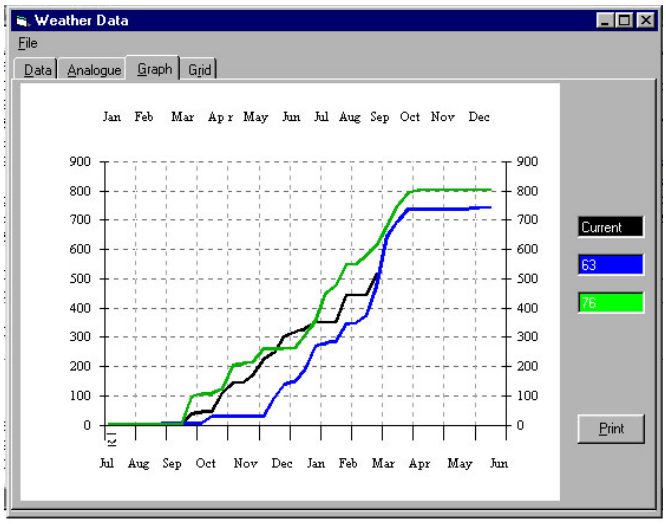


Figure 2. The current season in comparison with the two closest analogue years.

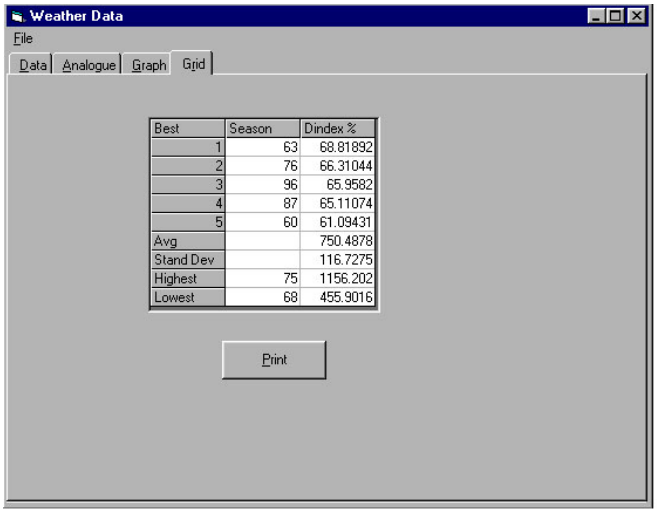


Figure 3. The five closest analogue years and weather statistics.

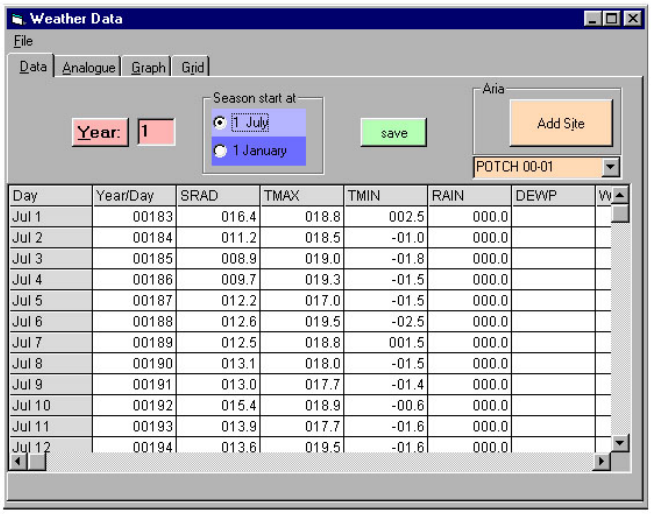


Figure 4. Spreadsheet type interface for data input, indicating both calendar date and Julian date.

Volume 3
Chapter 5

Graphical Display Program (EasyGrapher)

Jingyi Yang
E. C. (Ted) Huffman

Eastern Cereal and Oilseed Research Center, Ottawa, Canada
International Consortium for Agricultural Systems Applications

EasyGrapher v1.0

Help Manual

Sept. 30 2003

A Software for Graphic and Statistical Validation of DSSAT Outputs

Jingyi Yang and E. C. (Ted) Huffman
Correspondence address:
#4105, K. W. Neatby Building
Land Evaluation and modeling group,
Eastern Cereal and Oilseed Research Center
Agriculture and Agri-Food Canada
960 Carling Avenue, Ottawa, ON K1A 0C6
Canada

Contact by:

Jingyi Yang Tel: (613) 759-1412, Fax: (613) 759 1924; Email: yangj@agr.gc.ca;
yang_jingyi@hotmail.com
E.C. (Ted) Huffman: Tel: 613 759 1846: Fax: 613 759 1924; Email: huffmant@agr.gc.ca

Tables of Contents

1 About EasyGrapher

1.1 About EasyGrapher

EasyGrapher is a graphic display program designed for the DSSAT suite of models and for other crop simulation models such as, EPIC and N_ABLE. It allows users to manipulate time series and validation graphs for all simulation runs and it calculates validation statistics. EasyGrapher can be used to:

1. Create groups of simulation graphs (up to N runs by 12 variables, see Table 4 for details),
2. Display measured data in the simulation graphs,
3. Create validation graphs, e.g. graph simulated data (x) against measured data (y) for overall runs,
4. Calculate validation statistics such as RMSE, E, EF and paired-t, and display the statistics on the validation graph.

This menu will guide you using EasyGrapher smoothly, please contact our technical support if any problem happened when using the EasyGrapher.

1.2 Introduction-why EasyGrapher

Simulation models such as DSSAT, N_ABLE and EPIC are widely used in crop production studies, environmental assessment projects and soil and landscape evaluation exercises (Jagtap and Jones, 2002, Chung et al. 2001, Greenwood et al. 1996). When using these models to carry out an environmental assessment, one of the activities involves the calibration and validation of simulation results through recurrent comparisons of simulated and measured outputs (O'Leary and Connor, 1996, Asseng et al. 1998). However, interpretation and visualization of model outputs is time-consuming work mainly because agricultural simulation models tend to be diverse (outputs vary from one model to another) and many functions in the DOS operating systems and graphic outputs are not compatible with the windows environment. In order to graph these files using graphic tools, a user must open files, select data-blocks and then create graphs in a separate graphical software package. In addition, comparison of simulated and measured data requires the use of a statistical package to calculate statistics (Delgado et a. 2001, Jabro et al. 2001, Yang et al., 1999). This can be a daunting task when up to a thousand data files (outputs) need to be graphed and compared. EasyGrapher is designed to facilitate data visualization and statistical comparison, and to present simulation model outputs, especially for the DSSAT suite of models (Yang and Huffman, 2003).

1.3 Acknowledgement

We recognize financial support from the IROWCN-CRAM project sponsored by Agriculture and Agri-Food Canada (AAFC). We specially thank Gerrit Hoogenboom for the use of their experimental data for graph examples and for his many useful suggestions for EasyGrapher interface design, program testing and statistics. We also thank from AAFC: Reinder DeJong, Sam Gameda and Andy Bootsma for their review and comments, and Valerie Kirkwood for contributions to the help menu. Thanks also go to Chris duke for review and comments in Guelph University, Dewi Hartkamp for comments and for multi-years data testing in the Netherlands, and Oxana Uryaseva from University of Florida for help in VB programming in data type national conventions and localization issues.

1.4 Technical support

Land Quality Modelling Group
Eastern Cereal and Oilseed Research Centre
Agriculture and Agri-Food Canada

Dr Edward (Ted) Huffman
Study Leader
Land Use and Remote Sensing

#4097, K. W. Neatby Building
960 Carling Avenue, Ottawa, ON K1A 0C6
Tel: (613) 759 1846
Fax: 613 759 1924
Email: huffmant@agr.gc.ca
http://res.agr.ca/ecorc/

Dr Jingyi (Bill) Yang
Crop Modeller
Soil Science and Statistics

_____
#4105, K. W. Neatby Building
960 Carling Avenue, Ottawa, ON K1A 0C6
Tel: (613) 759 1412
Fax: (613) 759 1924
Email: yangj@agr.gc.ca: Email: yang_jingyi@hotmail.com
http://res.agr.ca/ecorc/

See author biography for details

1.5 Next version

Like any other software, EasyGrapher has its goals and limitations. The main constraints of EasyGrapher in the current version are:

(1) It can only graph Time-series outputs or similar data structure to the Time-series outputs, but not summary output, such as DSSAT summary.out, overview.out, etc.
(2) It cannot be used to draw bar graph, to map, or to draw surface contour graphs,
(3) It may not graph other models output well because of the differences in data formats and layout etc.

As discussed above, next steps may focus on overcoming limitations as shown above: Specifically, the next objective will be to;

(1) Graph DSSAT summary data,
(2) Graph other model output using EasyGrapher,
(3) Add more test statistics for the program, such as F-test, summary statistics etc.

2. Program design

2.1 Program design

EasyGrapher is written using Microsoft Visual BasicR (version 6) and automatically calls a MS ExcelR application (Microsoft Excel 1997/2000/2002) to perform a series of data handling and graphing tasks. Description of data flow, interfaces and routines follows. Details can be obtained from following topics.

Input files
Interface design
Outputs
System requirement

2.2 Input file

Three types of files from the DSSAT model are used as input files for EasyGrapher:

(1) Time-series outputs files (Table 1), DSSAT3.5 output files include growth.out, nitrogen.out, water.out, carbon.out and pest.out, DSSAT4 output files are listed in 'open file' topic.
(2) Measured data files, such as UFGA8101.SBT (Table 2) and
(3) DSSAT code label file, DATA.CDE, an example as shown in Table 3.

Note for input file:

EasyGrapher program is independent of DSSAT v4. Any missing of data in daily-step may affect the match between the simulated and the measured data when running EasyGrapher. We recommend that users save their DSSAT v4 output files as below:

(1) Always be in a daily time-steps in order to capture measured variables of any time-step interval.
(2) Headers in all RUNs must be identical in any single output file so that the EasyGrapher will properly label the Y-variables.

2.3 Interface design

Note: all interfaces in this menu refer to "Windows interface"

1. When running the EasyGrapher, this is the first interface which will appear on your screen.

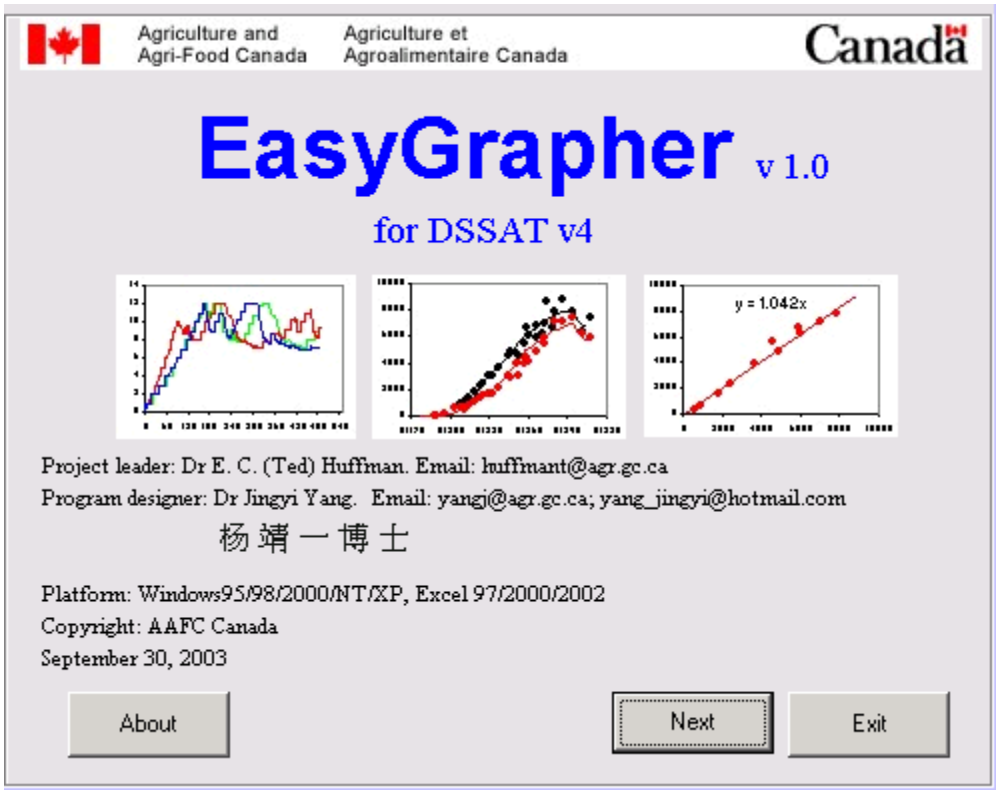


Fig. 1(a). The first interface of EasyGrapher.

After this, several interfaces are designed to run the program.

2. A common "open file" dialogue box appears to show DSSAT4 crop folders (by default); output files can be selected in the common dialogue box.

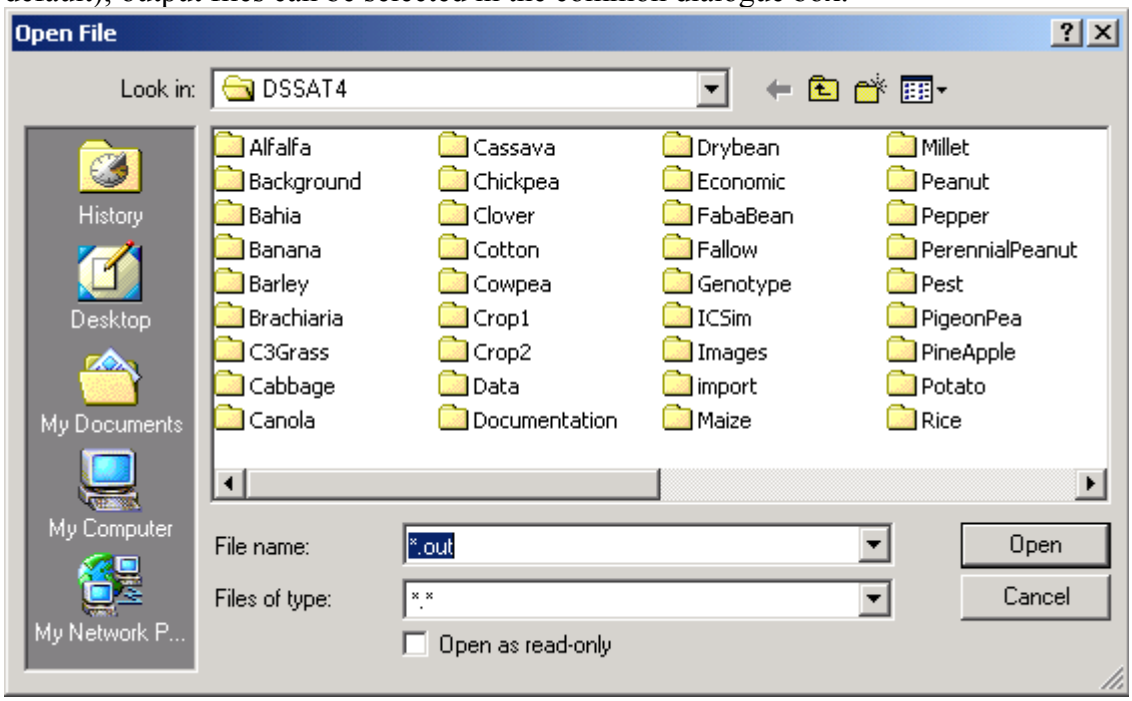


Fig. 1(b). Open File dialogue box of EasyGrapher.

3. An "input file" interface which allows users to change or confirm the graph file.

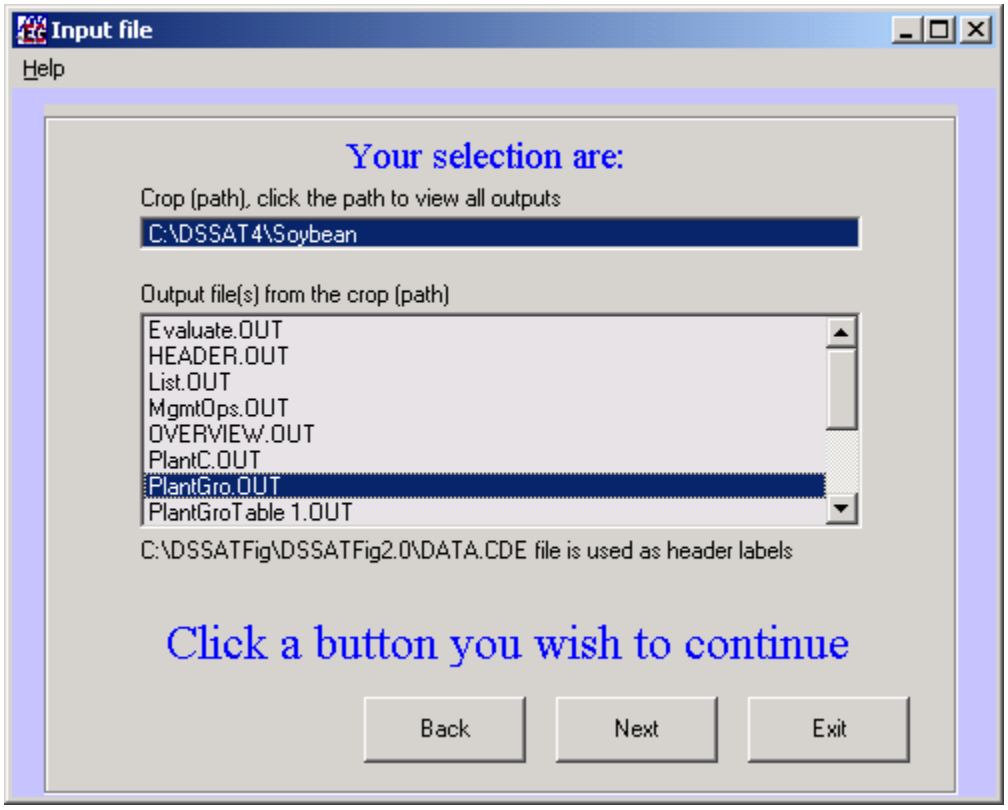


Fig. 1(c). Input file interface.

4. A "select variable" interface which allows users to select Y-variables, simulation RUNS and X-variables from three dialogue boxes.

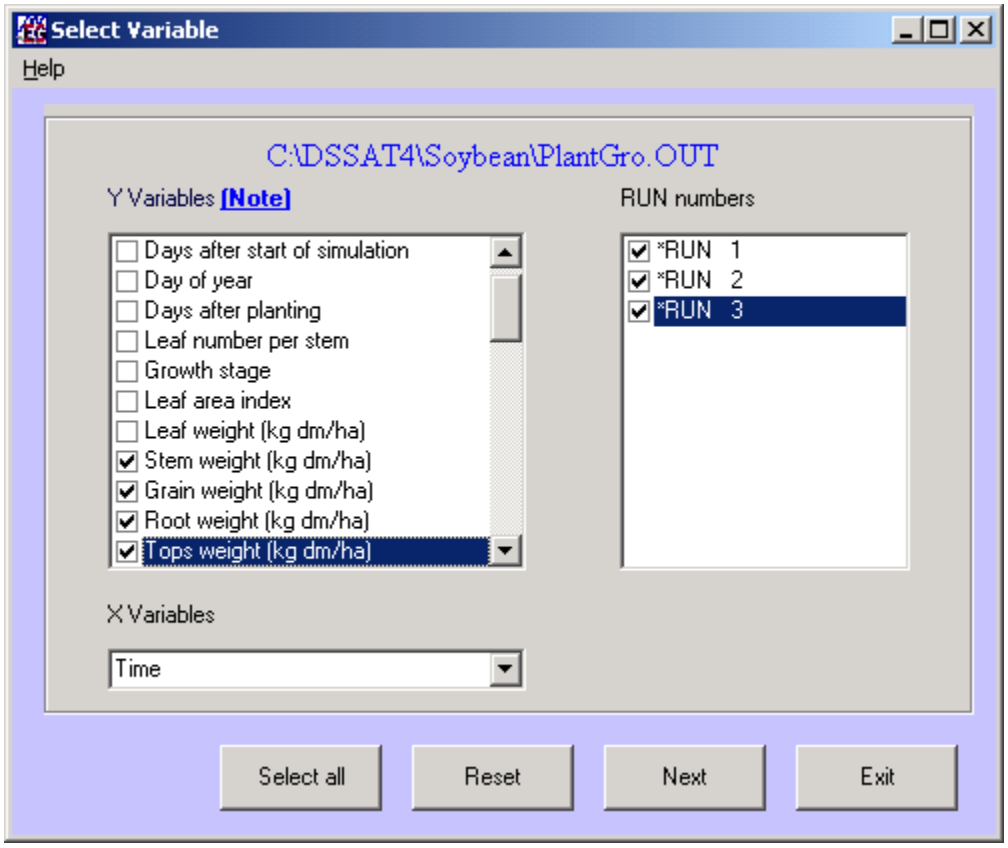


Fig. 1(d). Select variable interface.

5. When you select the "Next" button, a warning message appears to allow you select minimize the screen when the program is running. When you click the "Yes" button, the program runs in the background; when you click the "No" button, the program runs with an interface (Fig. 1(f)).

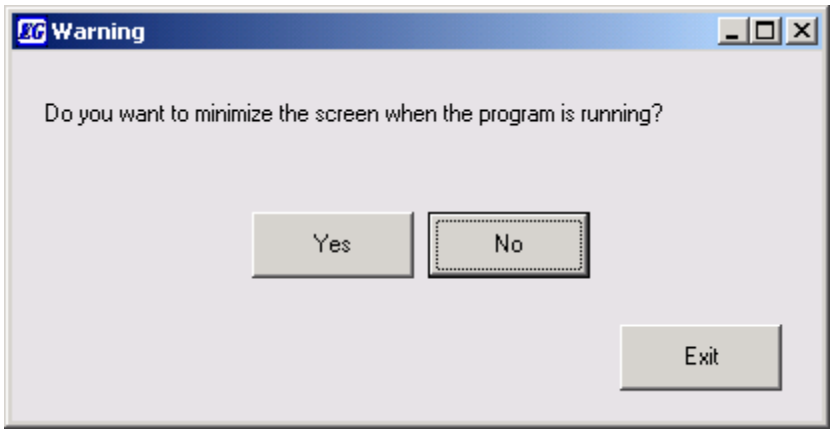


Fig. 1(e). Warning interface of EasyGrapher.

6. A "graph running" interface which includes a choice not to minimize the screen while the program is running, and

Fig. 1(f). Graph running interface with a progress bar.

7. A "graph summary" interface in which a run summary shows treatment name, total data numbers in each run and headers in each data file (Fig. not shown here).

With maximized screen in the "graph running" interface, a progress bar and information is presented. For example, when measured data is available, the information 'graph with measured data and validation statistics' appears, otherwise, 'graph without measured data' appears. It also presents information about the X variable you chose, e.g. 'data across years, crop age (days after starts of simulation) is used as X-variable'.

EasyGrapher assumes that the same headers/columns apply for all runs in one simulated output file, meaning that users cannot run EasyGrapher to select those runs whose headers are different from each other. Otherwise, a error may occur.

The maximum data headers allowed in this program is 100. In addition, online warnings appear on the screen when necessary and the main help menu can be invoked with 'help' button in each interface at any time or press the "F1" button from the keyboard.

Further information

To run EasyGrapher, follow the steps in Graph examples

Data flow between DSSAT and EasyGrapher can be seen from Chart 1 below.

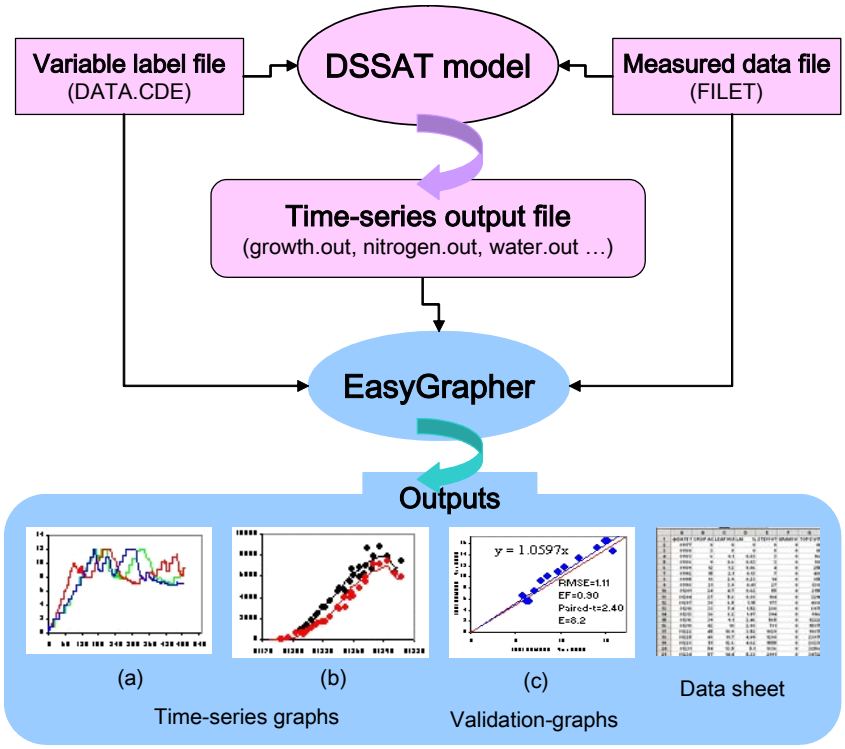


Chart 1. The data flow between DSSAT and EasyGrapher

2.4 Outputs

The main outputs from EasyGrapher are displayed in an Excel application in two data sheets and two graph sheets. They are;

(1) Time-series data,
(2) Validation-data,
(3) Time-series graph, read:
    a. Time-series-graph with simulated data;
    b. Time-series-graph with measured and simulated data
(4) Validation graph, see Validation (X-Y) graph

In addition, there are many separate inter-graph charts available for viewing.

2.5 System requirement

EasyGrapher has been tested for Windows 95/98/NT/2000//XP, and for Microsoft Excel 97/2000/2002, your computer should have following software installed.

(1) Windows 95/98/2000/NT/XP
(2) Microsoft Excel 97/2000/2002

In addition, when you install EasyGrapher from normal set up program, the supported files will be automatically installed to your computer. When you only copy EasyGrapher.exe to your DSSAT folder, you may need at least four support files to run the program;

(1) Mscomctl.ocx
(2) Tabctl32.ocx
(3) Comdlg32.ocx
(4) Msvbm60.dll

When you install the EasyGrapher, read readme.txt file for details.

2.6 EasyGrapher for other languages

EasyGrapher can be used in many languages supported by Windows, but the EasyGrapher interface in is only in English.

Theoretically, EasyGrapher v1.0 can be run in any language in both Windows and Microsoft Excel. Currently, EasyGrapher v1.0 has been tested in more than 30 languages in the Windows systems, but it has tested only in French and English Excel versions.

If EasyGrapher does not work properly in your local language of Microsoft Excel version, you need to set your MS office language to English, as below

To set Microsoft Office system to English (This is needed if your current language setting does not work)**.**

1. From **Start** menu of your computer->
2. **Programs**->
3. **Microsoft Office Tools**->
4. **Microsoft Office Language Setting**->
5. **User Interface tab**->
6. **English**
7. **O.K.**
8. **Re-Start** your computer

Be sure that your computer linked to at least one printer because EasyGrapher uses the printer's page set up, page orientation and zoom properties to view and print Excel graphs etc.

3 Data standard

3.1 Data standard

In general, EasyGrapher is designed to use the DSSAT data standards, but is not limit to it. This will be discussed in following topics.

Data structure
Search engine
Graph other model's data

3.2 Data structure

EasyGrapher is designed to use DSSAT (DSSAT3.5 and DSSAT4) data standards. When the program runs, DSSAT output file and measured data file are read and graphed, and another DSSAT file, DATA.CDE, is called for name labels. For all DSSAT data format and data standards, see Tsuji et al. (1994, V. 2-1). An example of a DSSAT growth output is shown in Table 1.

EasyGrapher is also suited for, but not limited to, DSSAT file name conventions. For example, EasyGrapher has the capability for any output files renamed with an extension name '.out' (e.g. GrowthMaize.out, Nitrogen-Barley.out and WaterWheat.out). For more information about DSSAT output files, read 'open file' topic for details.

When graphing other data, users can add their variable labels to the DATA.CDE file. For open DATA.CDE file, read 'open file' topic for details. For the format of graph other data, visit Graph other model's data topic for details.

3.3 Search engine

EasyGrapher uses only four pieces of information in the search engine for opening DSSAT files to graph; simulation run, input file name, crop, and data header entry (see words in bold letter (in blue color) in Table 1).

When an output file is opened, EasyGrapher first searches run numbers, e.g. '***RUN 1**' in line 2 of Table 1,

Then, searches for simulation input file name and crop, such as '**UFGA8101 SB**' in line 4 of Table 1.

After this, EasyGrapher searches for an ampersand, **@**, the first letter of the data headers line, assuming that the following letters in this line are headers, and each header ends with one or more spaces.

The next lines are assumed to be the data-block. The keywords serve as keys for the data entry, as we can see that ***RUN 1, UFGA8101, SB** and **@** only appear one time in the first simulation output Table 1.

3.4 Graph other model's data

Because EasyGrapher uses only four keywords for data entry, it can be used to graph similar output data from other models with similar data structures to that of DSSAT. In this case, users need to edit their output data according to the following steps:

In the simulated data output file:

(1) The first line of each run must be labelled ***RUN 1, *RUN** 2 etc. (1-3 spaces between RUN and 1, 2, 3).

(2) In a new line after the ***RUN** line, Add **EXPERIMENT**, file name (e.g. **UFGA8101**) and crop code (e.g. **SB**),

```
such as           ( EXPERIMENT     : UFGA8101 SB)
formatted as:     (SCCCCCCCCCCSSSSS:SCCCCIIIISCC)
where S = space, C=character and I=integer.
```

(3) Add **@** just before the first letter of the headers line of the data. There must be at least one space (not TAB) between headers. Similarly, at least one space (not TAB) between any two entries in each data lines. Note that users may save their data in a space-delimited format but not in a tab-delimited format.

(4) Data headers must be 3 to 5 characters long.

(5) The first column (field) must be a year in yyyy format (e.g. 1996) or year/date combination in yyddd format (e.g. 96120).

(6) The second column can be a data field labeled such as 'CDAY' (DSSAT v3.5) or 'DOY' (DSSAT v4), or any other header name you give which shows days starting from 1 at planting, numbered to the end of the simulation period.

(7) Add at least **one empty line** after each simulation (run).

(8) Save the file with an extension name 'out' (e.g. mydata.out).

In the measured data file (if one is used):

(9) EasyGrapher will search for a measured data file which matches the DSSAT naming convention given in (2) above, with an additional 'T' at the end of the file extension e.g. **UFGA8101.SBT** .

(10) Data headers must be 3-5 characters long. Use the same header names as in the simulated data file for similar data.

(11) Dates must be in yyyy or yyddd format.

(12) The first column must be labelled **TRN0** (treatment number). There must be the same number of treatments in the measured data file as there are runs in the

simulated data file. As in (3) above, the header line must start with an **@**, and have at least one space between headers.

An example of the measured data is given in Table 2. For more information, see Tsuji et al. (1994).

Edit the data.cde file:

(13) Edit the **data.cde** file to add any new header names and field definitions which are not standard for DSSAT output files. This applies to fields in both the simulation and measured data files.

4 Data access flexibility

4.1 Data access flexibility

In order to match DSSAT data standards and account for changes and flexibilities in DSSAT v4 outputs and measured data formats, the following automations apply:

Open file
Runs and Y-variables
Data records
Missing values
X-variables
Summaries

4.2 Open file

Crop folders, by default, are sub-folders in the DSSAT v4 directory. When the program runs, the DSSAT crop folders (e.g. C:\DSSAT4) are shown from a common dialogue box for users to select the 'crop' sub-directory and related output files.

EasyGrapher can open data files from any directory (e.g. from A:) through the common 'open file' dialogue box. Measured data files need to be located in the same directory as the output files (*.out file).

The DATA.CDE file can be located in C: D: E: or F:\DSSATab\ directory and its sub-directory, where a=0,… 9 and b = 0,… 9. For convenience, it can also be located in the same directory as EasyGrapher.exe. When DATA.CDE is located in any other directory, a warning message 'DATA.CDE is not found …, use CODE as header labels' appears on screen when running EasyGrapher.

EasyGrapher can produce graphs for time-series data, but not for outputs such as summary outputs or string outputs. Data that cannot be graphed are protected when opening a file.

Files (or renames) for which graphs can be produced using EasyGrapher

ETPhot.out	MgmtOps.out	PlantC.out	PlantGro.out
PlantN.out	SoilC.out	SoilN.out	SoilTemp.out
SoilWat.out	Weather.out		

Files for which graphs cannot be produced using EasyGrapher

soilNbal.out	plantNbal.out	soilWatbal.out
Evaluate.out	Header.out	List.out

Measured.out	Overview.out	Plantera.out
Planters.out	Plantert..out	Plantev.out
Plantrem.out	Plantres.out	Plantsum.out
Summary.out	Work.out	

4.3 Runs and Y-variables

As discussed above, EasyGrapher will search all runs throughout the output data, and automatically display runs, X and Y-variables in three dialogue boxes in the "select variable" window for user selection.

A maximum of 12 Y-variables can be selected each time when EasyGrapher is run, and the maximum 'RUNS' can be selected depending on the Y-variables being selected, see Table 4 for details.

4.4 Data records

When reading data, EasyGrapher counts data headers in the selected run as names of Y-variables. After it finishes reading the header line, the program reads a data line, and sets each data to its header from left to right. Headers will be used as Y-variable names during graphing.

Data records: There is no limitation for data rows/records
Maximum data columns: The maximum data columns/fields are set to 100 in the current version.

4.5 Missing values

The program assumes that ".", "-99.0", "-999", "****", "*****.**" are all missing values. When the missing value is read, it is set to a null string.

4.6 X-variables

When simulations cross years, the program automatically sets 'crop age' (the second column in the output data) as the X variable. Because EasyGrapher uses Excel's graph tool, and Excel interprets a 'yyddd' data format in a decimal system, this may lead to a mis-scaled X-axis in the graph.

For example, in a multi-year dataset, the X-axis may be from 92301 … 92365, 93001 … to 93300. In this case, the data is interpreted as 92301 … 92365, 62366, … 92999, 93001 … 93300, as a consequence, the X-axis in the graph is mis-scaled.

The following cases are considered as time across year:

(1) Simulated data across years,
(2) Measured data across years,
(3) Both simulated and the measured data across years,
(4) RUNS are from different years (run 1 is from 1993 and run 2 is 1995).

When runs are from different years, EasyGrapher searches for the same year in the measured files as the simulated data in that run.

There is no year limitation.

4.7 Summaries

The program copies graphs automatically to the Excel sheets. If measured data is graphed, there are four Excel sheets for displaying graph data, graphs and statistics:

(1) Time-series-data sheet,
(2) XY-validation-data-sheet that contains statistics,
(3) Time-graphs,
(4) Validation-graphs.

For more information, see graph examples.

5 Graph examples

To run EasyGrapher, follow the steps 1 to 6 below;

(1) When running EasyGrapher, the first interface will appear on your screen (see Fig. 1(a) in Interface design):
(2) In the "Open File" dialogue box, click the crop folder and file that you wish to graph (see Fig. 1(b) in Interface design).
(3) In the "Input File" interface, when input file is selected (highlighted) in , click the "Next" button to continue (see Fig 1(c) in Interface design).
(4) In the "Select Variable" interface, select Y variables from Y variable list box, select X variable from X variable list box, and select RUN numbers from "RUN numbers" list box (see Fig. 1(d) in Interface design). When X,Y variables and RUN numbers are selected, click the "Next" button to continue.
(5) In the "Warning" interface (see Fig. 1(e) in Interface design), a warning message appears to allow you select minimize the screen when the program is running. Click the "No" button to continue.
(6) A "program running" interface with a progress bar will appear on your screen to record your running progress (see Fig. 1(f) in Interface design).

When finished, Excel graphs are opened for you, See Fig. 2, Fig. 3 and Fig. 4 for graph examples in followng topic.

In the following examples, EasyGrapher is used to graph several outputs from DSSAT v4 in order to show its graphing potentials and statistical power. The data used in the following graph examples are from the DSSAT v4 directory, and the reference is given in each graph.

Time-series-graph with simulated data
Time-series-graph with measured and simulated data
Validation-graph

5.2 Time-graph with simulated data

EasyGrapher can graph the simulated data without the measured data. When this is the case, the program can be run fast and efficiently, meaning that more runs can be drawn in each graph or more Y-variables can be selected each time when the program is run (Table 4). If you do not have the measured data file in your crop directory, EasyGrapher only graphs the simulated data. If you have the measured data in the crop directory and you prefer to graph only the simulated data, simply move the measured data file (TFILE) out of the crop directory, then run the EasyGrapher.

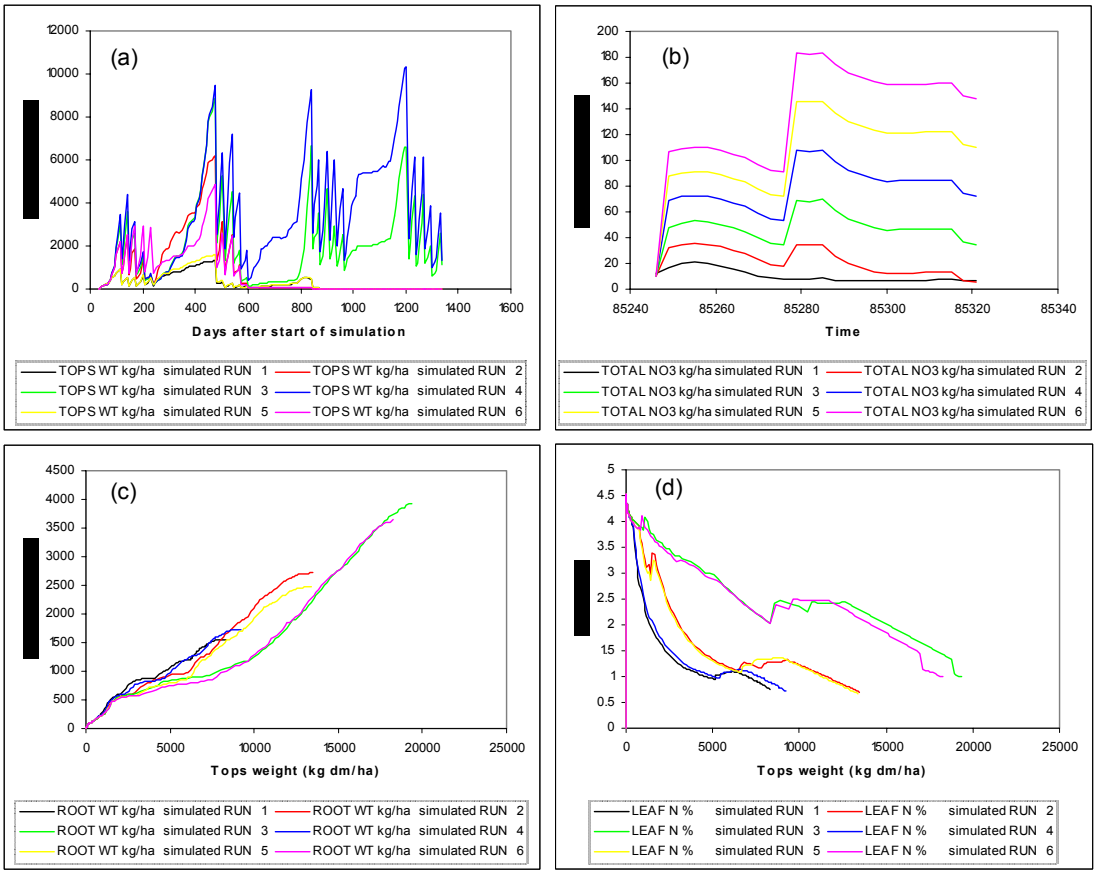


Fig. 2. Time-series graphs with simulated data;
(a) Top dry weight against crop age (4 year bahiagrass experiment, ALTH5601);
(b) total NO3-N against time (rice experiment, DTSP8502);
(c) root dry weight against top dry weight (maize experiment, IBWA8301)
(d) leaf % N trends against top (above round) dry weight (maize experiment, IBWA8301).

To graph the example Figures

To graph Fig. 2(a)

(1) Follow step 1 and 2 of Graph examples. Select a graph file in the example folder:
EasyGrapher\Example\Bahia\PlantGro.out.
If you run ALTH5601 experiment with DSSAT v4. the file is in:
C:\...DSSAT4\Bahia\PlantGro.out.

(2) Follow step 3 and 4 of Graph examples. In the "Select Variable" Interface, select;
Y="Top weight";
RUN=1,2,3,4,5,6;
X="Time".

(3) Click "No" button on "warning message" in step 5 of Graph examples. The program then runs with an progress bar.

(4) When finished, Fig. 2(a) appears on your screen in the Excel sheet.

To graph Fig. 2(b)

(1) Follow step 1 and 2 of Graph examples. Select a graph file in the example folder:
EasyGrapher\Example\Rice\SoilN.out.
If you run DTSP8502 experiment with DSSAT v4, the file is in:
C:\...DSSAT4\Rice\SoilN.out.
(2) Follow step 3 and 4 of Graph examples. In the "Select Variable" Interface, select;
Y="Soil NO3-N";
RUN=1,2,3,4,5,6;
X="Time".
(3) Click "No" button on "warning message" in step 5 of Graph examples. The program then runs with an progress bar.
(4) When finished, Fig. 2(b) appears on your screen in the Excel sheet.

To graph Fig. 2(c)

(1) Follow step 1 and 2 of Graph examples. Select a graph file in the example folder:
EasyGrapher\Example\Maize\PlantGro.out.
If you run IBWA8301 experiment with DSSAT v4, the file is in:
C:\...DSSAT4\Maize\PlantGro.out.
(2) Follow step 3 and 4 of Graph examples. In the "Select Variable" Interface, select;
Y="Root weight";
RUN=1,2,3,4,5,6;
X="Top weight".
(3) Click "No" button on "warning message" in step 5 of Graph examples. The program then runs with an progress bar.
(4) When finished, Fig. 2(c) appears on your screen in the Excel sheet.

To graph Fig. 2(d)

(1) Follow step 1 and 2 of Graph examples. Select a graph file in the example folder:
EasyGrapher\Example\Maize\PlantGro.out.
If you run IBWA8301 experiment with DSSAT v4, the file is in:
C:\...DSSAT4\Maize\PlantGro.out.
(2) Follow step 3 and 4 of Graph examples. In the "Select Variable" Interface, select;
Y="Leaf N%";
RUN=1,2,3,4,5,6;
X="Top Weight".
(3) Click "No" button on "warning message" in step 5 of Graph examples. The program then runs with an progress bar.
(4) When finished, Fig. 2(d) appears on your screen in the Excel sheet.

5.3 Time-graph with measured data

EasyGrapher can graph the measured data with the simulated data in one Time-series graph. When the program runs, it displays in the run time window information such as "graph with the measured data".

EasyGrapher searches for each measured data and sorts the measured data according to the simulated data headers.

To graph measured data with simulated data, prepare the measured data file (TFILE) in the same directory as the simulated output file (*.out), then run EasyGrapher.

If there is no measured data in the measured data file to match the simulated data, the measured data are set to a null string.

This type of graph (Fig. 3) is widely used in the model validation process as it provides a visual tool to demonstrate the goodness of fit between simulated and measured data in the Time-series. Outliers (unusual data) can be discerned easily from the graph.

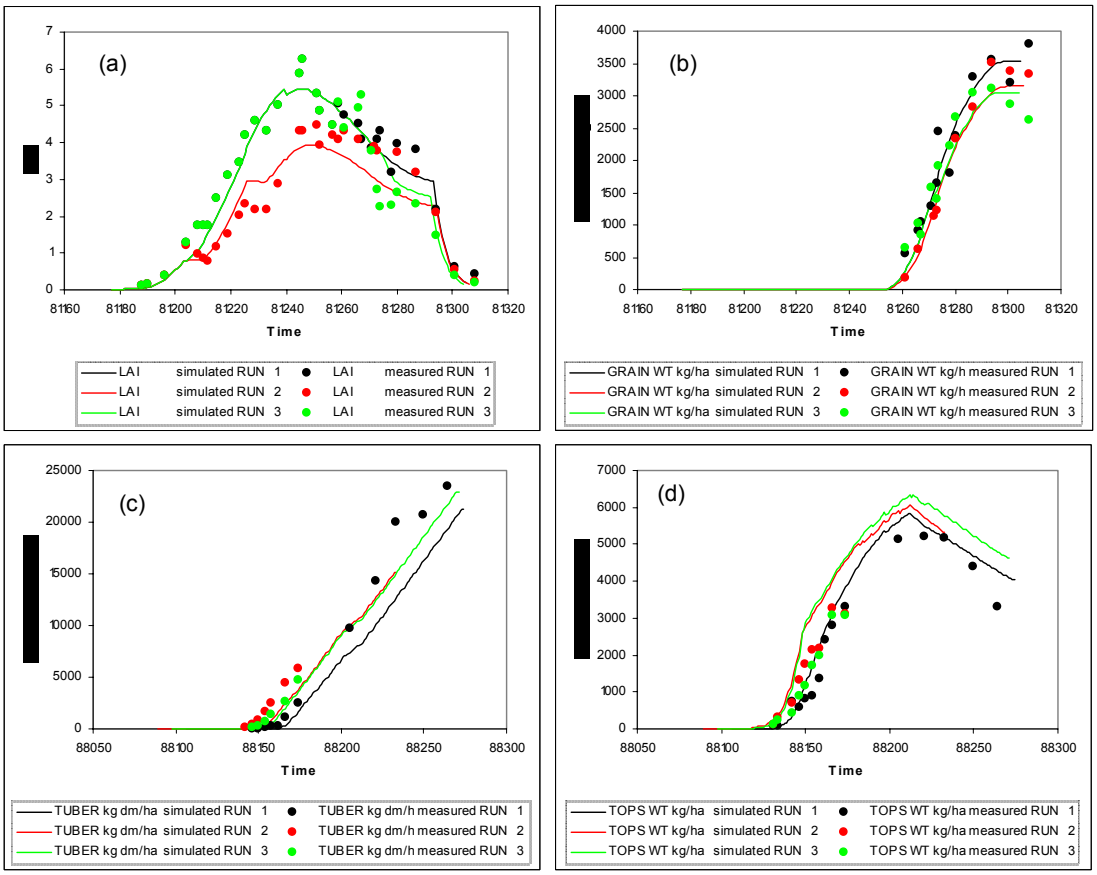


Fig. 3. Time-series graph with measured data;
(a) and Fig. (b) are leaf area index and grain weight against time (soybean experiment, UFGA8101)
(c) and Fig. (d) are tube dry weight and tops (above ground) dry weight against time (potato experiment, OSBO8801).

Follow the steps below to produce the example Figures
To graph Fig. 3(a, b) and Fig. 4(a,b)

(1) Follow step 1 and 2 of Graph examples. Select a graph file in the EasyGrapher example folder:
    EasyGrapher\Example\Soybean\PlantGro.out.
    If you run UFGA8101 experiment with DSSAT v4, the file is in:
    C:\...DSSAT4\Soybean\PlantGro.out.
(2) Follow step 3 and 4 of Graph examples. In the "Select Variable" Interface, select;
    Y="Leaf area index", and Y="Grain weight"
    RUN=1,2,3;
    X="Time"
(3) Click "No" button on "warning message" in step 5 of Graph examples. The program then runs with an progress bar.
(4) When finished, Excel graphs appear on your screen.
    In the "Time-graph" sheet of Excel, Fig. 3(a) and Fig. 3(b) appear;
    In the "XY-graph" sheet of Excel, Fig. 4(a) and Fig. 4(b) appear.

To graph Fig. 3(c,d) and Fig. 4(c,d)

(1) Follow step 1 and 2 of Graph examples. Select a graph file in the EasyGrapher example folder:
    EasyGrapher\Example\Potato\PlantGro.out.
    If you run OSBO8801 experiment with DSSAT v4, the file is in:
    C:\...DSSAT4\Potato\PlantGro.out.
(2) Follow step 3 and 4 of Graph examples. In the "Select Variable" Interface, select;
    Y="Tuber weight", and Y="Top weight"
    RUN=1,2,3;
    X="Time"
(3) Click "No" button on "warning message" in step 5 of Graph examples. The program then runs with an progress bar.
(4) When finished, Excel graph appear on your screen.
    In the "Time-graph" sheet of Excel, Fig. 3(c) and Fig. 3(d) appear;
    In the "XY-graph" sheet of Excel, Fig. 4(c) and Fig. 4(d) appear.

5.4 Validation (X-Y) graph

EasyGrapher can graph measured data against simulated data. This kind of graph is known as a validation-graph or XY-graph (Fig. 4). Validation graphs are useful visual tools to illustrate the goodness of fit between the simulated and the measured data in general. The validation graph consists of:

(1) One 1:1 line
(2) One linear regression (y = bx)
(3) Four statistics.

EasyGrapher is designed to calculate a variety of statistics, making it easy to test a model with modern statistics within minutes. Visit validation statistics for more information.

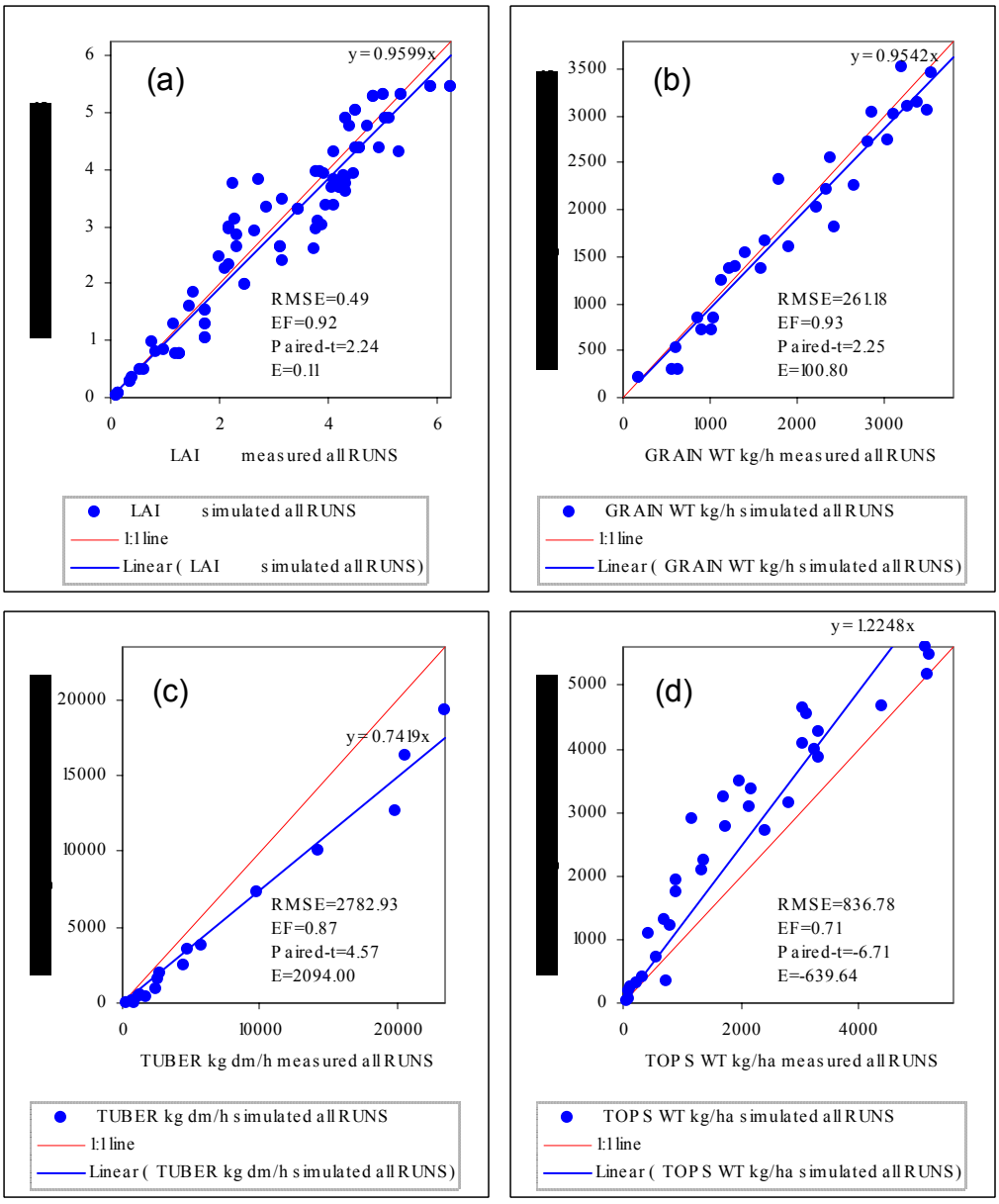


Fig. 4. Validation-graph and statistics;

(a) and (b) are the measured leaf area index and grain weight against the simulated data (soybean experiment, UFGA8101);

(c) and (d) are the measured tube and tops (above ground) dry weights against the simulated data (potato experiment, OSBO8801).

To graph Fig. 4

Follow "To graph Fig. 3(a, b) and Fig. 4(a,b) and "To graph Fig. 3(a, b) and Fig. 4(a,b)" in " Time-series-graph with measured and simulated data".

6 Validation statistics

6.1 Validation statistics

Validation statistics are an important feature of EasyGrapher. Two test statistics and six difference statistics are calculated by EasyGrapher using both measured and simulated data. Statistics used in EasyGrapher are discussed below.

Test statistics
Difference statistics
More information about statistics

6.2 Test statistics

Test statistics used in the EasyGrapher are linear regression and paired-t as shown in the Eq. (1) and Eq. (2).

$$y_i = bx_i + e_i \qquad (1)$$

where y_i and x_i are the measured and simulated values for the ith measurement, i = 1,2, … n, e_i represents random error in the ith data in Eq. (1). When testing with Eq. (1), y is regressed on x and the null hypothesis is tested with H0: $b = 1$.

Paired-t statistic is often used to test the null hypothesis H0: H_0: $\bar{d}$ ($\bar{y}_i$ - $\bar{x}_i$) = 0 (Reckhow et al., 1990);

$$paired\ t = \bar{d} / s_{\bar{d}} \qquad (2)$$

$$s_{\bar{d}} = \sqrt{\sum_{i=1}^{n} (d_i - \bar{d})^2 / n(n-1)}$$

where $\bar{d}$ is the mean value of the difference variable, and $s_{\bar{d}}$ is the standard error of the mean.

Test statistics follows three assumptions, for more information, see our publication (Yang et al. 2000).

6.3 Difference statistics

To avoid violation of assumptions behind test statistics, difference statistics have been developed in recent decades so as to test d_i (y_i - x_i) directly without considering data assumptions and transformations. A detailed discussion of the difference statistics was given in our previous paper (Yang et al. 2000) and in other publications (e.g. Kabat et al.

1995). Here we employ six of statistics: Root of Mean Square of Error (*RMSE*), Mean Error (*E* or *ME*), Mean absolute Error (*MAE*), Forecasting Efficiency (*EF*), modified Forecasting Efficiency (EF_1), and Coefficient of error (*C*) as defined below:

$$RMSE = \sqrt{\Sigma(y_i - x_i)^2 / n} \quad (3)$$

$$E = \Sigma(y_i - x_i) / n \quad (4)$$

$$MAE = \Sigma \mid y_i - x_i \mid / n \quad (5)$$

$$EF = 1 - \Sigma(y_i - x_i)^2 / \Sigma(y_i - \bar{y})^2 \quad (6)$$

$$EF_1 = 1 - \Sigma \mid y_i - x_i \mid / \Sigma \mid y_i - \bar{y} \mid \quad (7)$$

$$C = \Sigma \mid y_i - x_i \mid / n / \bar{y} = MAE / \bar{y} \quad (8)$$

Eq. (1) and Eq. (8) were coded into EasyGrapher as a statistics routine, and the calculated statistics and inter-variables are displayed in the validation data-sheet of the Excel. Four statistics are displayed in the validation-graphs

6.4 More information about statistics

The *RMSE* was also named Root Mean Square of Deviation, *RMSD*, by other researchers (Probert et al. 1998, Kobayashi and Salam, 2000). It has been increasingly used to validate simulation models, including the DSSAT model (Kobayashi and Salam 2000, Iramk et al. 2000, Hoogenboom et al. 1999). The mean error *E* was also named *ME* by others (e.g. Kabat et al. 1995), or named *MD* (Garrison et al. 1999). It is an indicator of whether the model predictions tend to over- or under-estimate measured data (Addiscott and Whitmore, 1987). Because positive and negative errors can negate each other in the *E* value, *MAE* was used as an alternative statistic by Willmott et al. (1985). *RMSE, E* and *MAE* all take on units of $d_i = y_i - x_i$ and are mainly used as measures of accuracy to compare the output of the same variables.

EF is a relative measure of error used by many authors (Loague and Green, 1991; Loague and Freeze, 1985). Like R^2, *EF* can be used to compare the accuracy of model outputs for different variables because it has no unit. $EF = 1$ if $y_i = x_i$ and $EF < 1$ for any realistic simulation. $EF < 0$ if the model predicted values are worse than simply using the measured mean of y_i. $EF \neq R^2$ because values of R^2 are $0 \leq R^2 \leq 1$. Values of R^2 change with different linear models, such as $y = a + bx$, $y = bx$ and $y = x$. The *EF* statistic is actually a statistic to test the goodness of fit of the model $y = x$. *EF* statistic is considered to be sensitive to outliers because it uses a quadratic residual function (Klepper and Rouse, 1991). Yang et al. (2000) defined another function EF_1 as an alternatives to *EF*, by replacing the sum of squares of difference with the sum of absolute differences as