Available online at www.sciencedirect.com
SCIENCE DIRECT
Computers and Electronics in Agriculture 45 (2004) 125–132
Computers and electronics in agriculture
www.elsevier.com/locate/compag

Application note

EasyGrapher: software for graphical and statistical validation of DSSAT outputs

J.Y. Yang*, E.C. (Ted) Huffman

Eastern Cereal and Oilseed Research Center, Agriculture and Agri-Food Canada, #4105, K.W. Neatby Building, 960 Carling Avenue, Ottawa, Ont., Canada K1A 0C6

Received 28 May 2003; received in revised form 2 May 2004; accepted 13 June 2004

Abstract

EasyGrapher is a graphical display and statistical validation program designed for the decision support system for agro-technology transfer (DSSAT) suite of models. EasyGrapher can expedite validation of DSSAT output, which normally requires considerable time and effort to export output data into external statistical packages. It allows users to create validation graphs, displaying simulated data against ground-truth data and calculate validation statistics such as root mean square error, mean error, forecasting efficiency and paired *t*-tests. It operates under Microsoft Windows, is free of charge, and is available as part of a distribution toolkit with DSSAT v4.0.

Keywords: EasyGrapher; DSSAT; Data visualization; Validation statistics

1. Introduction

Simulation models are widely used in crop production studies, environmental assessment projects, and soil and landscape evaluation exercises. However, interpretation and visualization of the diverse range of agricultural model outputs is time-consuming. In addition, comparison of simulated and measured data usually requires the use of an external software program to calculatestatistics (Garrison et al., 1999; Greenwood et al., 1996; Yang et al.,

* Corresponding author. Tel.: +1 613 759 1412; fax: +1 613 759 1924.
E-mail address: yangj@agr.gc.ca (J.Y. Yang).

doi:10.1016/j.compag.2004.06.006

2000). This can be a daunting task when up to a thousand output data files need to be graphed and compared. EasyGrapher is designed to facilitate data visualization and statistical comparison, and to present simulation model output, especially for the DSSAT (Hoogenboom et al., 2004) suite of models. The aim of this paper is to: (i) introduce EasyGrapher; (ii) show how EasyGrapher presents DSSAT output; and (iii) discuss validation statistics and explain their power.

2. Program design

EasyGrapher is written using Microsoft Visual Basic[1] (v6) and has been tested for Windows 95/98/2000/NT/XP and for Microsoft Excel[1] 97/2000/2002. It functions through a group of data handling routines, which automatically call a Microsoft Excel application to perform the data graphing tasks. DSSAT output is read by a read-data routine, the program opens an Excel application and the data are placed in the Excel sheet. A measured-data file is then opened and the data are read to the spreadsheet. A statistical routine is invoked to compare the simulated data with the measured data and to calculate statistics. When the statistical calculations are finished, graphing routines are called sequentially to draw different types of graphs according to variables selected in the "select variable" dialog (see below). As a final step, validation statistics are read to Excel, and graphs are reformatted and copied to new sheets for viewing and printing (see below for details).

User interface dialogs are available to open files, select graph variables and view output. The main dialog is "select variable" and it allows users to select *X*- and *Y*-axis variables and specify simulation runs (Fig. 1). Warnings appear on screen when necessary and a help menu can be invoked via a menu item or by pressing "F1" on the keyboard. The help menu and system support files can be automatically installed from the EasyGrapher set up program.

3. Data standards

3.1. Data input

EasyGrapher is designed to use DSSAT data standards. When the program runs, three types of files from the DSSAT model are used as input files for EasyGrapher: (1) the DSSAT output file, (2) a measured data file, and (3) a file that associates data header labels with output variable names (DATA.CDE). For example, LWAD is a header label for "leaf weight (kg dm/ha)". For DSSAT data format and data standards see (Tsuji et al., 1994). If the DATA.CDE has not been created or is not found, data-file header labels are used as graph labels.

EasyGrapher keys in on four pieces of information to graph DSSAT data files. The four keywords are underlined in the following excerpt from DSSAT v4 output.

[1] The use of trade names does not constitute endorsement.

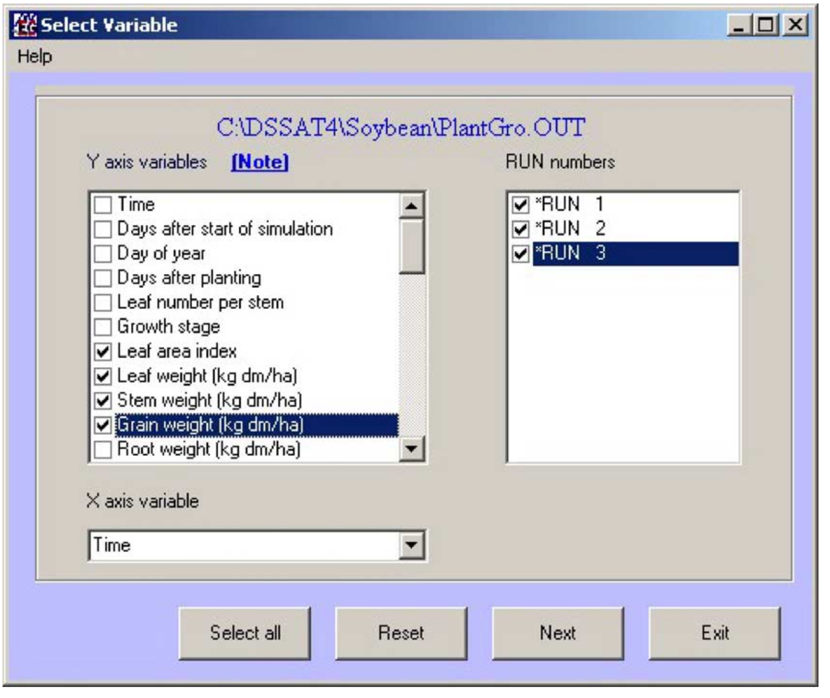

Fig. 1. The “Select Variable” dialog of EasyGrapher. *Y*-axis variables that can be selected are a function of RUN numbers, with a maximum of 12 variables. The *X*-axis can be set to any output variable, and the default time is in yyddd format. There is no limit on the selection of RUNs.

```
*GROWTH ASPECTS OUTPUT FILE

*RUN    1          :  IRRIGATED
 MODEL             : CRGRO020 - SOYBEAN
 EXPERIMENT        : UFGA7801 SB     UFGA7801SB BRAGG, IRRIGATED & NON-IRRIGATED
 TREATMENT  1      :  IRRIGATED

@YEAR DOY   DAS    DAP  L#SD  GSTD  LAID  LWAD  SWAD  GWAD  RWAD  CWAD  G#AD   GWGD
 1978 166     0      0   0.0     0  0.00     0     0     0     0     0     0    0.0
```

EasyGrapher first searches for a run number (e.g. ‘*RUN 1’), then for a measured data file name (e.g. ‘UFGA7801 SB’), and then for an @ symbol. The @ symbol is the first character of the data-headers line in DSSAT, and EasyGrapher assumes that the following terms in this line are headers and that each header ends with one or more spaces. The search routine then assumes that lines following the header line contain the data.

In order to match DSSAT v4 data standards and to account for variations in DSSAT output and measured data formats, the following protocols apply:

3.1.1. Open file

When EasyGrapher begins, available DSSAT crop folders are shown as defaults in a dialogue box. These are used to select the ‘CROP’ sub-folder and the related output files (*.out). Measured data files must reside in the same directory as the output files.

3.1.2. Selecting variables for graphing

EasyGrapher displays all runs and output variables (headers) in the ‘select variable’ interface (Fig. 1), and any output variable can be used for the *X*- and *Y*-axis in the graph. For each run of EasyGrapher, only one variable can be selected for the *X*-axis (the default is time), while up to 12 variables can be selected for the *Y*-axis (in this case, 12 graphs drawn). There is no limitation on the selection of runs.

3.1.3. Date records and missing values

When reading data, EasyGrapher interprets data headers in the selected run as the names of *Y*-variables. After reading the header line, the program reads each data line, and sets each value to its corresponding *Y*-variable from left to right. There is no limitation on the number of data rows/records, but the maximum number of header/data columns is set at 100 in the current version. The program assumes that “.”, “−99.0”, “−999”, “****”, “*****.**” are all missing values. When a missing value is encountered, it is set to a null string.

4. Graph examples

EasyGrapher can graph measured data and simulated data in one time-course graph. Under this type of application, EasyGrapher searches for measured data and sorts it according to the headers for the simulated data. This type of graph is widely used in model validation processes, as it provides a means of visualizing the goodness of fit between simulated and measured data over time. Outliers (unusual data) can be identified easily from the graph. Leaf area index and grain weight of soybean from a 1981 experiment are displayed in Fig. 2a and b.

EasyGrapher can graph measured data against simulated data, known as a validation-graph or *XY*-graph (Fig. 3). Validation graphs are useful to illustrate the goodness of fit between simulated and measured data in general. Validation graphs between the measured and simulated leaf area index and grain weight of soybean from a 1981 experiment are displayed in Fig. 3a and b.

Robust testing and validation of simulation models requires the calculation of a variety of test statistics. Preparing data for statistical packages and running those analyses can often be time consuming. Validation statistics are an important feature of EasyGrapher, and are designed to perform a variety of calculations that are used frequently by modelers. Generally, when graphing simulated data against measured data, an overall validation-graph is created for each variable, a 1:1 line is added and a regression line, a paired-*t* and several difference statistics are calculated.

Difference statistics have been developed to test $d_i = y_i - x_i$, (where d_i is the difference between y_i and x_i, the measured and simulated values for the *i*th measurement) directly without considering data assumptions and transformations. A detailed discussion of difference statistics can be found elsewhere (e.g., Yang et al., 2000); here we focus on six of them: root of mean squared error (*RMSE*), mean error (*E* or *ME*), mean of absolute error

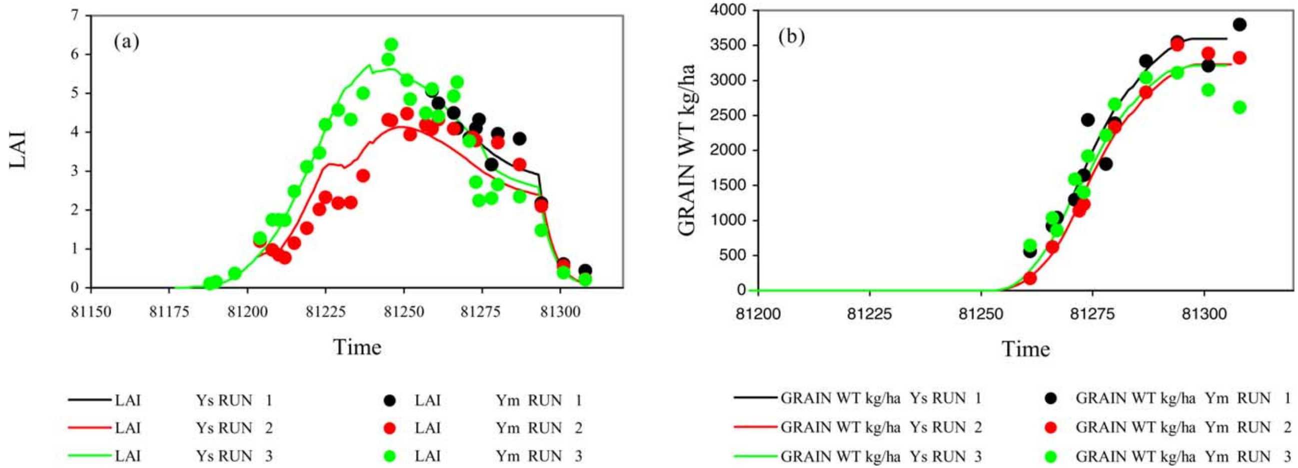

Fig. 2. Time-course graphs with measured and simulated data for (a) leaf area index (LAI) and (b) grain weight (GRAIN WT).

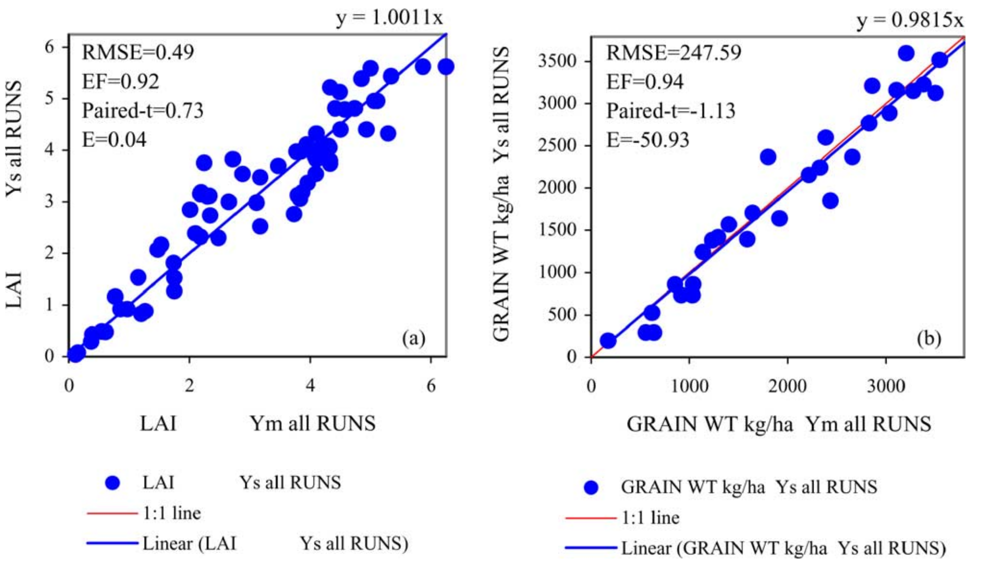

Fig. 3. Validation-graphs and statistics for (a) leaf area index (LAI) and (b) grain weight (GRAIN WT).

(*MAE*), forecasting efficiency (*EF*), modified forecasting efficiency (EF_1), and coefficient of error (*C*). These statistics, calculated by EasyGrapher, are defined as:

$$RMSE = \sqrt{\frac{\sum (y_i - x_i)^2}{n}} \tag{1}$$

$$E = \frac{\sum (y_i - x_i)}{n} \tag{2}$$

$$MAE = \frac{\sum |y_i - x_i|}{n} \tag{3}$$

$$EF = 1 - \frac{\sum (y_i - x_i)^2}{\sum (y_i - \bar{y})^2} \tag{4}$$

$$EF_1 = 1 - \frac{\sum |y_i - x_i|}{\sum |y_i - \bar{y}|} \tag{5}$$

$$C = \frac{\sum |y_i - x_i|}{n/\bar{y}} = \frac{MAE}{\bar{y}} \tag{6}$$

Each of these statistics has particular uses and limitations. *RMSE* (also referred to as root mean square of deviation) has been used to validate simulation models, including DSSAT

(Kobayashi and Salam, 2000; Hoogenboom et al., 1999). Mean error (also termed mean difference) (Garrison et al., 1999) is used to identify whether model predictions tend to over or under-estimate results compared to measured data. Since positive and negative errors can negate each other in the mean error, the mean of absolute error (*MAE*) is often used as an alternative (e.g. Willmott et al., 1985). *RMSE*, *E*, and *MAE* are used as measures of accuracy for simulated outputs. *EF* is a relative measure of error used by many researchers (e.g. Loague and Green, 1991; Loague and Freeze, 1985). Yang et al. (2000) compared a group of statistics for model validation, and defined another function, $EF_{1,}$ as an alternative to *EF* because the sum of squares in *EF* tends to enlarge the real differences, where $|EF_1| \leq |EF|$. *C* is another relative average measure of absolute difference which is expressed as a proportion of the mean of the measured variable (Klepper and Rouse, 1991). Fig. 3 shows *XY*-graphs of experimental results with five validation statistics.

5. General discussion

EasyGrapher v1.0 is designed to graph time-series outputs or data structures similar to time-series outputs, but not summary outputs such as DSSAT's summary.out or overview.out, and it cannot be used to draw bar graphs or surface contour graphs. However, EasyGrapher can be used for graphical and statistical validation of DSSAT output within minutes and can graph simulated data without measured data. In the latter case, it can show differences in state-variables. It provides DSSAT users with a handy tool for model validation and statistical analysis. For these reasons it has been accepted as a tool of the DSSAT v4.0 program and will be distributed with DSSAT v4.0 free of charge. The program has also been tested satisfactorily on output from DSSAT v3.5. It will be available in the near future for downloading from the 'Tool Kit' of ICASA's website at: http://www.icasa.net/dssat/.

EasyGrapher can also be used to graph output data from other models, if the model source code can be accessed in order to modify output data files (or the output files can be post-processed). EasyGrapher was used successfully to graph output from a Canadian Agricultural Nitrogen Budget (CANB) model (Yang et al., 2003) and from a plant nitrogen balance simulation model in Germany (Kersebaum and Beblik, 2001). Theoretically, EasyGrapher can be used to graph any dataset from field or lab experiments as long as the data are edited or formatted as DSSAT output. Detailed instructions on this topic can be found in the help menu, which accompanies the program. In addition, the program has been successfully tested in a number of languages that Windows and MS Excel support. Currently, it can be obtained by e-mail from the corresponding author, and it is recommended that users update EasyGrapher regularly.

Acknowledgements

We wish to acknowledge the financial support of the National Agri-Environmental Health Analysis and Reporting (NAHARP) program of Agriculture and Agri-Food Canada (AAFC). We especially thank Gerrit Hoogenboom of the University of Georgia, for the use of his team's experimental data for the examples in this paper, and for his many useful suggestions on interface design, program testing and statistics. We also thank AAFC specialists, Andy Bootsma and Sam Gameda, for their review and comments, and Valerie Kirkwood for assistance with the help menu. Thanks also goes to Chris Duke at Guelph University for review and comments, and Oxana Uryaseva from University of Florida for help in VB programming in data type national conventions and localization issues.

References

Garrison, M.V., Batchelor, W.D., Kanwar, R.S., Ritchie, J.T., 1999. Evaluation of the CERES-maize water and nitrogen balances under tile-drained conditions. Agric. Syst. 62, 189–200.

Greenwood, D.J., Rahn, C.R., Draycott, A., Vaidyanathan, L.V., Paterson, C., 1996. Modelling and measurement of the effects of fertilizer-N and crop residue incorporation on N-dynamics in vegetable cropping. Soil Use Manage 12, 13–24.

Hoogenboom, G., Wilkens, P.W., Tsuji, G.Y. (Eds.), 1999. DSSAT v3, vol. 4. University of Hawaii, Honolulu, HI.

Hoogenboom, G., Jones, J.W., Wilkens, P.W., Porter, C.H., Batchelor, W.D., Hunt, L.A., Boote, K.J., Singh, U., Uryasev, O., Bowen, W.T., Gijsman, A.J., du Toit, A., White J.W., Tsuji, G.Y., 2004. Decision support system for agrotechnology transfer version 4.0 [CD-ROM]. University of Hawaii, Honolulu, HI.

Kersebaum, K.C., Beblik, A.J., 2001. In: Shaffer, M., et al. (Ed.), Performance of a nitrogen dynamics model applied to evaluate agricultural management practices. Modeling carbon and nitrogen dynamics for soil management. Lewis Publishers, Boca Raton, pp. 549–569.

Klepper, O., Rouse, D.I., 1991. A procedure to reduce parameter uncertainty for complex models by comparison with real system output illustrated on a potato growth model. Agric. Syst. 36, 375–395.

Kobayashi, K., Salam, M.U., 2000. Comparing simulated and measured values using mean squared deviation and its components. J. Agron. 92, 345–352.

Loague, K., Green, R.E., 1991. Statistical and graphical methods for evaluating solute transport models: overview and application. J. Contam. Hydrol. 7, 51–73.

Loague, K.M., Freeze, R.A., 1985. A comparison of rainfall-runoff modelling techniques on small upland catchments. Water Resour. Res. 21, 329–348.

Tsuji, G.Y., Jones, J.W., Balas, S. (Eds.), 1994. DSSAT v3, vol. 2. University of Hawaii, Honolulu, HI.

Willmott, C.J., Ackleson, S.G., Davis, R.E., Feddema, J.J., Klink, K.M., Legates, D.R., O'Donnell, J., Rowe, C.M., 1985. Statistics for the evaluation and comparison of models. J. Geophys. Res. 90, 8995–9005.

Yang, J., Greenwood, D.J., Rowell, D.L., Wadsworth, G.A., Burns, I.G., 2000. Statistical methods for evaluating a crop nitrogen simulation model, N_ABLE. Agric. Syst. 647 (1), 37–53.

Yang, J.Y., Huffman, E.C. (Ted), De Jong, R., 2003. EasyGrapher v2.0: software for graphic and statistical analysis of agri-environmental model outputs. Poster in: 2003 ASA-CSSA-SSSA Annual Meeting, Denver, CO.