EasyGrapher: Software for Data Visualization and Statistical Evaluation of DSSAT Cropping System Model and the CANB Model

J. Y. Yang, C. F. Drury, J. M. Yang, Z. T. Li, and G. Hoogenboom

Abstract—Crop simulation models are process-based and can provide very detailed outputs of the simulated crop and soil processes. However, analyzing the outputs is challenging. EasyGrapher (EG) is a graphical and statistical software program designed for the DSSAT v4.5 crop models, and it also supports graphing Canadian Agricultural Nitrogen Budget (CANB) v3.0 model. EG allows users to manipulate hundreds of graphs within minutes and calculates evaluation statistics. Specifically, EG v4.5 can support 16 outputs of DSSAT v4.5 to: (1) create time-series graphs and display measured data in the simulation graphs; (2) graph simulated data against measured data and (3) calculate evaluation statistics (RMSE, E, EF and d).

EG v4.5 can support outputs of CANB v3.0 to create bar plots for spatial-series graphs in various scales. The EG program has the potential to carry out graphic and statistical tasks for other models. Graphic and statistical evaluation examples were illustrated using field dataset collected from our experiments in Canada and China, as well as dataset resided in the DSSAT software.

Index Terms—Cropping system models, data visualization, decision support system for agro-technology transfer (DSSAT), EasyGrapher, process-based simulation, statistical evaluation.

I. INTRODUCTION

The Decision Support System for Agro-technology Transfer (DSSAT) is a computer software that combines many crop models. It uses the same soil and water models to simulate crop growth, soil C/N dynamics [1], and soil water balance under different climate zones and management strategies [2], [3]. In 1998, the crop simulation models integrated to the DSSAT v3.0 model were documented in a book [2, 3]. In 1998, the crop simulation models integrated to the DSSAT v3.0 model were documented in a book (20 research articles) and were regarded as the principal references in DSSAT model training and application [4]. DSSAT v3.5 introduced the Cropping System Model (CSM) that combined that various individual crop model into one main model [2]. The Century-based soil SOM model was also integrated to DSSAT [5], [6]. The soil water balance model was updated by separating the soil and plant evaporation and plant root uptake into a SPAM model [2], where the CSM covered more than 25 crops mainly from legumes (soybean, peanut, dry bean, chickpea, cowpea, faba bean, tomato, Macuna, Brachiaria and Bahiagrass) and cereals (CERES-Maize, CERES-Wheat, CERES-Rice, SubStor-Potato etc.). DSSAT v4.0 was released with a Microsoft Windows version. Now the current version is DSSAT v4.5 [3]. During the last 20 years, DSSAT models have been increasingly used worldwide. A review regarding the use of DSSAT to model cropping systems was prepared by Sarkar [7] and used more than 100 references. Recently, the update of DSSAT-Century soil model in DSSAT v4.5 was reported [6] and tillage simulation was integrated to CSM in order to simulate the effect of tillage to crop residues on soil [8]. There are other upgrades on DSSAT input databases, such as using the World Inventory of Soil Emission Potentials (WISE) database to parameter DSSAT soil input [9], using different methods for estimating cultivar coefficients [3], [10], [11], and addition of the Weather Man software developed to format DSSAT weather data [12].

EasyGrapher (EG) is a software package designed by Agriculture and Agri-Food Canada for graphical and statistical evaluation of DSSAT outputs [13], [14]. EG calculates several statistics, such as residual mean square error (RMSE) and index of agreement (d) to evaluate the simulated results with v4.5 outputs, including various plant and soil outputs (i.e. nitrogen, carbon, water, temperature) and summary outputs.

In addition to graph DSSAT outputs, EG also support to graph other model’s outputs [14]. In 2003, EasyGrapher v2.0 was designed to support graphing for Canadian Agricultural Nitrogen Budget (CANB) model [15]. The CANB model is a landscape model developed at 1:1M scale of agricultural soil polygon in Canada [16]. The CANB program calculated two agro-environmental indicators: residual soil nitrogen (RSN) and Indicator of Risk of Water Contamination by N (IROWC-N) from 3000 soil polygons annually starting from 1981. A daily water dynamic model, WSMB, was integrated into the CANB v3.0 model to calculate water drainage on each polygon, and then estimate N lost and N concentration from the agricultural soil profile. The RSN, N lost and IROWC-N values were scaled up to eco-regional, provincial and Canadian scales for national agri-environmental health analysis and report [17].

The objective of this paper is to introduce the EasyGrapher v4.5 program, including its interface design, graphic types, statistics and potential applications in graphical and statistical evaluation of the DSSAT model and the CANB model.
II. PROGRAM DESIGN

A. Compiler and Working Environment

EasyGrapher is written using Microsoft Visual Basic (.NET). It functions by automatically calling a Microsoft Excel application to perform a series of data handling and graphing tasks. It works under a MS Windows environment and is compatible with Windows XP/Vista/7/8 and Microsoft Excel XP/2007/2010. EG also support most language user interfaces under MS Windows and MS Office. In the following section we provide a description of the data flow and the interface.

B. DSSAT Graph File

Three types of files from the DSSAT model are used as input files for EasyGrapher. 1) EG supports 11 DSSAT time-series outputs plus two summary outputs (Table I). 2) Measured data for model performance evaluation, such as measured soybean data file UFGA8101.SBT. 3) A code label file (DATA.CDE) containing variable code and descriptions [4], [18]. The main outputs from EasyGrapher are displayed in an Excel data sheets and graph sheets (see graph section for details). Outputs (i.e., Overview. out) that cannot be graphed are protected when opening a file.

TABLE I: LISTS OF DSSAT V4.X* GRAPH FILES FROM SOYBEAN EXPERIMENT (UFGA8101.SBX)

<table>
<thead>
<tr>
<th>DSSAT Output</th>
<th>Description</th>
<th>Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>PlantN.out</td>
<td>Plant N</td>
<td>19</td>
</tr>
<tr>
<td>PlantGro.out</td>
<td>Plant growth</td>
<td>45</td>
</tr>
<tr>
<td>PlantC.out</td>
<td>Plant C</td>
<td>17</td>
</tr>
<tr>
<td>SoilWat.out</td>
<td>Soil water</td>
<td>23</td>
</tr>
<tr>
<td>SoilTemp.out</td>
<td>Soil temperature</td>
<td>12</td>
</tr>
<tr>
<td>SoilOrg.out</td>
<td>Soil OM</td>
<td>16</td>
</tr>
<tr>
<td>SoilN.out</td>
<td>soil N</td>
<td>33</td>
</tr>
<tr>
<td>Weather.out</td>
<td>Weather</td>
<td>17</td>
</tr>
<tr>
<td>MgmtOps.out</td>
<td>Management</td>
<td>10</td>
</tr>
<tr>
<td>ET.out</td>
<td>Evaporation</td>
<td>28</td>
</tr>
<tr>
<td>Mulch.out</td>
<td>Mulch option</td>
<td>6</td>
</tr>
<tr>
<td>Summary.out</td>
<td>Summary</td>
<td>65</td>
</tr>
<tr>
<td>Evaluate.out</td>
<td>Evaluate</td>
<td>25</td>
</tr>
</tbody>
</table>

* refers different version of DSSAT 4, such as DSSAT v4.0, 4.5, 4.6 etc.

C. Program Interface

The main interfaces designed for EasyGrapher include: 1) a main startup screen with initial start options (Fig. 1a); additional functionality is added at run-time (i.e. re-graph capability for multiple runs of EasyGrapher). 2) An "Open Output File" dialogue box displays DSSAT45 crop folders (by default); output files can be selected in the common dialogue box (Fig. 1b). 3) A "Select Data to Display" interface allows users to select Y-axis variables, X-axis variables and Treatment numbers from three dialogue boxes (Fig. 1c). 4) A "Progress" interface appears on the screen while the program is running (not shown). MS Excel workbook graphs appear on the screen when EasyGrapher ends, see GRAPH DISPLAYS section for details.

The EasyGrapher help manual contains contents, index and search capabilities similar to all other help software. It is easy to be invoked by clicking the F1 key from the keyboard when the EG runs.

III. EVALUATION STATISTICS

Test statistics, such as testing $R^2$ significance, and $\alpha=0$, etc.
\[ \beta = 1, y_i = \alpha + \beta x_i + \varepsilon, \] require the error term \( \varepsilon \) follow three assumptions (normality, equal variance, and independence) while in crop growth measurement, though sample data may violate one of the assumptions. For example, the measured biomass from planting to harvest often violate equal variance. To avoid violation of assumptions behind test statistics, deviation statistics have been developed in recent decades so as to test deviation \( d_i = y_i - x_i \) directly without considering data assumptions and transformations. A detailed discussion of the deviation statistics was given by [19]-[21]. In the current version of EG more than 10 deviation statistics were employed, including: Root Mean Square Error (RMSE), Mean Error (E or ME), Modeling Efficiency (EF) [22] and index of agreement (\( d \)) [23] as defined in Fig. 1.

\[
RMSE = \sqrt{\frac{\sum (y_i - x_i)^2}{n}} \quad (1)
\]

\[
E = \frac{\sum (y_i - x_i)}{n} \quad (2)
\]

\[
EF = 1 - \frac{\sum (y_i - x_i)^2}{\sum (\bar{y} - \bar{x})^2} \quad (3)
\]

\[
d = 1 - \frac{\sum (y_i - x_i)^2}{\sum (|y_i - \bar{y}| + |x_i - \bar{x}|)^2} \quad (4)
\]

All statistics are calculated by EasyGrapher program and saved in the evaluation data-sheet of the Excel, and four statistics (Eqs. 1-4) are displayed in the in the evaluation-graphs (see Fig. 4).

IV. GRAPH DISPLAYS

A. Time-Series Graph

![Time-Series Graph](image)

Fig. 2. (a) Time-series top dry weight against crop age (from a 50 year long term field experiment, Ontario, Canada) using DSSAT sequence analysis; (b) leaf area index against crop age (from a 7 cultivars by 50 year maize experiment (DTCM0301), Changmai, Thailand) using DSSAT seasonal analysis.

EasyGrapher uses only four pieces of information in the search engine for opening DSSAT files to graph time-series output, such as plant growth outputs. EG first searches for treatment (run) numbers, e.g. "**RUN 1"", then searches for the "EXPERIMENT" line, and then finds a measured data file name, e.g. "UFGA7801.B". After this, EG searches for a "@" symbol which is the first character of the data headers line. It assumes that the letters following in this line are headers, and that each header ends with one or more spaces. The next lines are assumed to be the data-block. If there is an empty line following the data-block, EG assumes the data-block ends, and then it continues to search for "**RUN 2" and so on. Please view "output file" (Table II), or read [14] for details. Time-series graphs can be displayed with/without measured data. In case of no measured data (no FILET), EG only plots simulated data against time (see Fig. 2). When the measured data files are available, EG plots both simulated data (line) and measured data (point) on the graphs (see Fig. 3).

B. Evaluation Graph

![Evaluation Graph](image)

Fig. 3. Time-series graph with measured data: (a) top dry weight against crop age (DSSAT soybean experiment, UFGA8001); (b) soil mineral N (kg N ha\(^{-1}\)) against crop age (maize, CHUN0801, China).

TABLE III: MEASURED DATA (UFGA801.BT) FROM DSSAT INPUT FILE

<table>
<thead>
<tr>
<th>EXPERIMENT</th>
<th>DASK</th>
<th>DASB</th>
<th>DASC</th>
<th>DASM</th>
<th>DAT</th>
<th>DTEN</th>
<th>DTEN</th>
<th>DTEN</th>
<th>DTEN</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>-9</td>
<td>-9</td>
<td>11</td>
<td>0</td>
<td>18</td>
<td>254</td>
<td>1366</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>-9</td>
<td>-9</td>
<td>63</td>
<td>1.27</td>
<td>0</td>
<td>254</td>
<td>1366</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>-9</td>
<td>-9</td>
<td>11</td>
<td>0.44</td>
<td>-9</td>
<td>2307</td>
<td></td>
</tr>
</tbody>
</table>

1 @TRN0 refers to a header line in the experimental data file; the numbers in the first column match the *RUN numbers in the output files shown in Table II.

The model evaluation graphs plot the measured data on the X-axis against the simulated data on the Y-axis. The evaluation graphs can be shown by the measured file,
FILET (i.e., UFGA8101.SBT) (Table III), containing time course measured data during the simulation, mainly from start to end of the simulation. In most cases, it covers crop and soil samples from planting to harvest. The evaluation graphs for USA soybean experiment (UFGA8101.SBX) and for China maize experiment (CHUN0801.MZX) were shown in Fig. 4. Four evaluation statistics were also shown in the evaluation graph (see Fig. 4).

C. Summary Graph

The EasyGrapher v4.5 can also graph DSSAT summary outputs such as Summary. out. DSSAT summary file includes only final outputs of key variables varying from growth outputs (growing stage, harvest days, treatment levels, plant biomass, leaf, stem, shell, yield at maturity, root weights and plant N uptake etc) and soil N components (soil NH₃-N, NO₃-N and mineral N in each layer, N leaching, mineral N, nitrification and denitrification etc.). EG opens summary output (summary. out) and produces intermediate summary and evaluation formatted files, readily available for EG graphs. The Summary graphs are displayed by a line-graph with scatter points (see Fig. 5).

D. Graph Other Model Using EasyGrapher

Since 2003, EasyGrapherv2.0 has been designed to support graphing for Canadian Agricultural Nitrogen Budget (CANB) v1.0 to v3.0 model [15]. The CANB model and its functions were given in the introduction section, and here we illustrate some bar plot graphs of the CANB v3.0 output using the updated EG v4.5.

The base level of spatial series outputs from the CANB v3.0 model was from 1:1M scale soil polygons at Soil Landscapes of Canada (SLC) (i.e., about 3000 SLCs). Therefore, bar plot is a suitable graph type for the SLC output in the CANB v.30 model. When the spatial data was scaled up to province, the weighted averages of residual soil nitrogen (RSN) and other outputs were calculated with a spatial reference of province. The bar plot, therefore, can be used to graph sensitivity or scenario analysis outputs by province with different levels of parameter changes (see Fig. 6).

V. SUMMARY

The strength of the EG program is its speed. For example, EGv4.5 can graph 13 DSSAT outputs (Table I) within 15 minutes to graph more than 300 time-graphs, plus many evaluation graphs and calculated evaluation statistics. Secondly, EG has power to graph more than 200 treatments in seasonal simulation (Fig. 2b), and thirdly, EG allows users to plot any graph between output and input parameters and these graphs are very important for sensitivity analysis of output to input parameters. In addition, EG program has power to graph other models output, such as the CANB model, with a small integration work required (i.e., a subroutine to support other models).

Like other software, EG has its weak point. For example, there is no internal integration of the I/O data between EG v4.5 and DSSAT v4.5. When using EasyGrapher to graph DSSAT outputs, we recommend that the users set up daily output equal to 1 day (i.e. setup FROOT=1 in X file) so that EasyGrapher can easily match the measured outputs without missing. The purpose of the EG graphs is to allow users to do quick graphical tasks and then check the relationships among output variables. Precise or professional graphs should be re-drawn or modified if users wish to add them for publication. In the future, EG should be updated to support many other models’ outputs.

ACKNOWLEDGMENT

We acknowledge the financial support from Agriculture...
and Agri-Food Canada (AAFC). We thank the DSSAT team for providing support in testing EasyGrapher during the last 10 years and for the use of their team’s experimental data for graphing; especially thank P.W. Wilkens for his support to test and integrate EasyGrapher v4.5 program with the DSSAT v4.x software.

REFERENCES


