

Guidelines For Installing and Running GLUE For Cultivar and Ecotype Coefficient Estimation

DSSAT Version 4.8.5

Last updated: December 1, 2024

Overview

The GLUE (Generalized Likelihood Uncertainty Estimation) program is used to estimate genotype-specific coefficients for the DSSAT crop models (Jones et al., 2003; Hoogenboom et al., 2019). 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 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 using a recommended number of model runs for calibration. What happens then is that the GLUE program will run twice, once for determining the phenology coefficients and another for determining the 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 cultivar (CUL) file (e.g., MZCER048.CUL or SBGRO048.CUL, etc.) and, optionally, into the ecotype (ECO) file (e.g., MZCER048.ECO or SBGRO048.ECO, 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 FileA in DSSAT, which are typically taken once per season (mostly at harvest), are used in the estimation process. Time series data recorded in FileT is not utilized by GLUE for calibrating genetic coefficients.

There are several assumptions that may have important effects on the resulting parameters. First are the prior distributions of coefficients, which are stored in each individual crop cultivar and ecotype file. 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 files named "MeasurementVariance_All.csv" and others. 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.

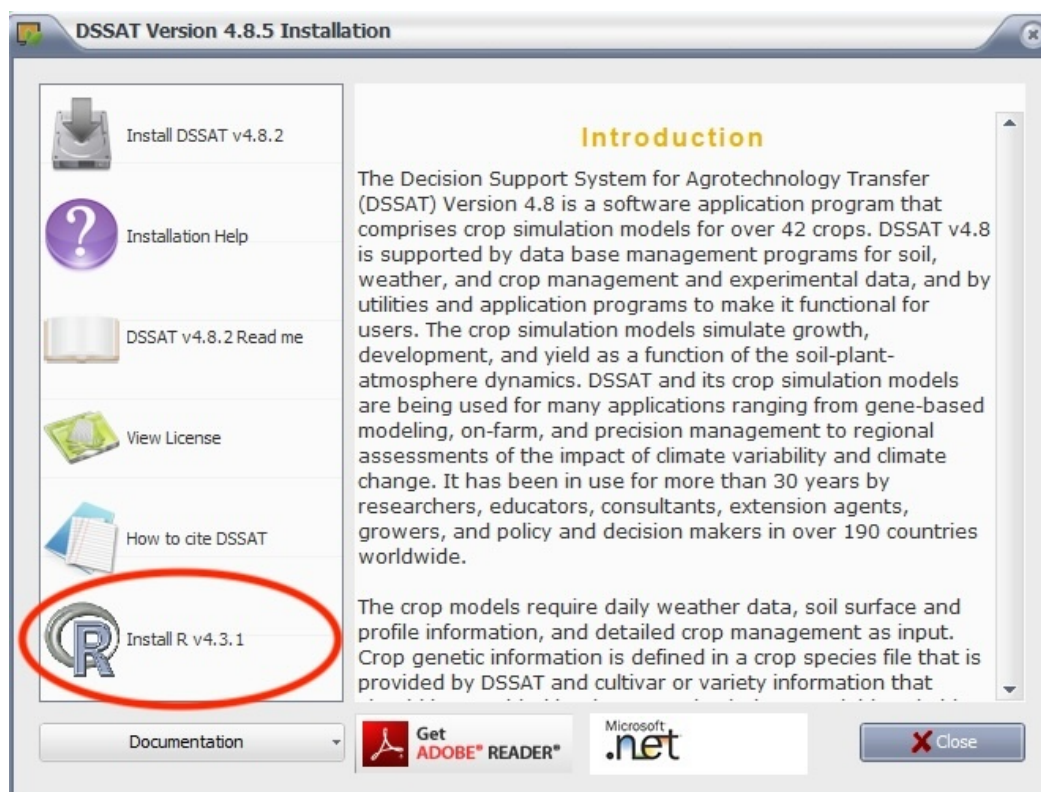
There are other cautions that users should be aware of. For example, results from an estimation process provide conditional estimates of coefficients. That means that the coefficients are the best set given the measurements that were used, but the coefficients also depend on the set of observations used 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 three tools in DSSAT for estimating cultivar coefficients for the different crops, alongside GENCALC and Time-Series cultivar coefficient Estimator (Hunt et

al., 1993; He et al., 2010; Memic et al., 2021). Each tool has its own strengths and limitations. GLUE, as a global search technique, can explore the entire parameter space, minimizing the risk of converging to local optima. However, parameter estimation using GLUE has been often correlated to extensive computations that requires a substantial amount of time. To address this, the GLUE version integrated in DSSAT Version 4.8.5 incorporates the Generalized Likelihood Uncertainty Estimation Parallelized (GLUEP), a R-written program that uses the GLUE technique and parallel computing to significantly reduce calibration time (Ferreira et al., 2024). Moreover, GLUE can autonomously generate a set of estimated coefficients along with their uncertainty estimates, without user intervention. 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.



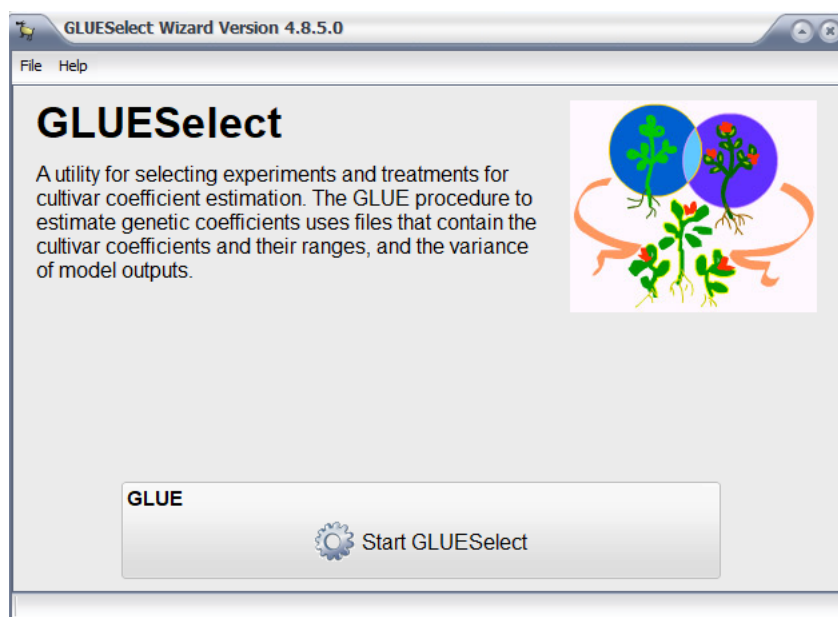
3. Use of the GLUE Procedure to Estimate Genetic Coefficients

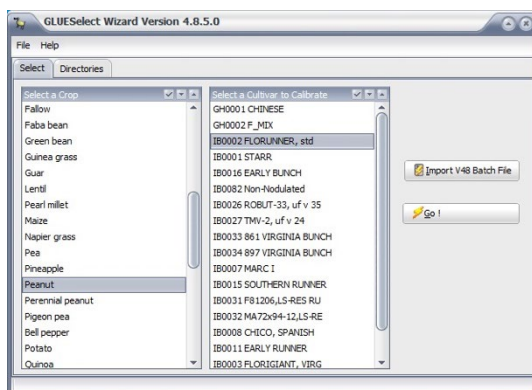
3.1 Setting conditions for GLUE to estimate coefficients

The GlueSelect program, developed by Paul Wilkens (IFDC) in Delphi, is a user-friendly graphical interface introduced in DSSAT Version 4.5. It serves as a tool to help users interact seamlessly with the GLUE program. 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 v4.7 standards). 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 defines the uncertainty associate with data observation for most crops is the “MeasurementVariance_All.csv” (see Appendix A). This file is stored in the directory “C:\DSSAT48\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.

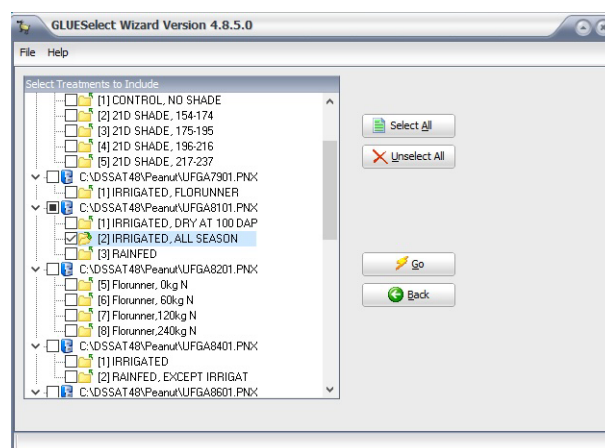
The GLUE program is integrated into the DSSAT48 shell, and the user runs the GlueSelect program from the DSSAT Tools menu to start the process as shown below.



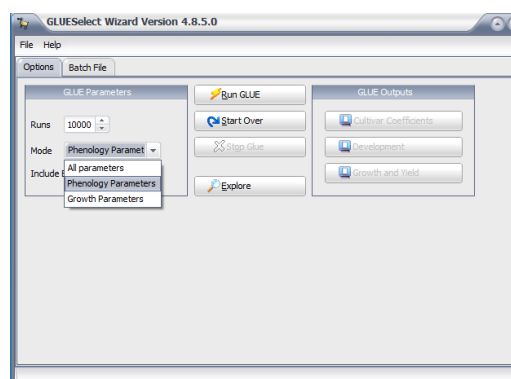


growing season was selected from a single peanut experiment. A batch file will be created in the DSSAT48\GLWork directory (Appendix B), and these treatments will be simulated, using the GLUE method to estimate the coefficients that give the maximum likelihood for both phenology and growth measurements.

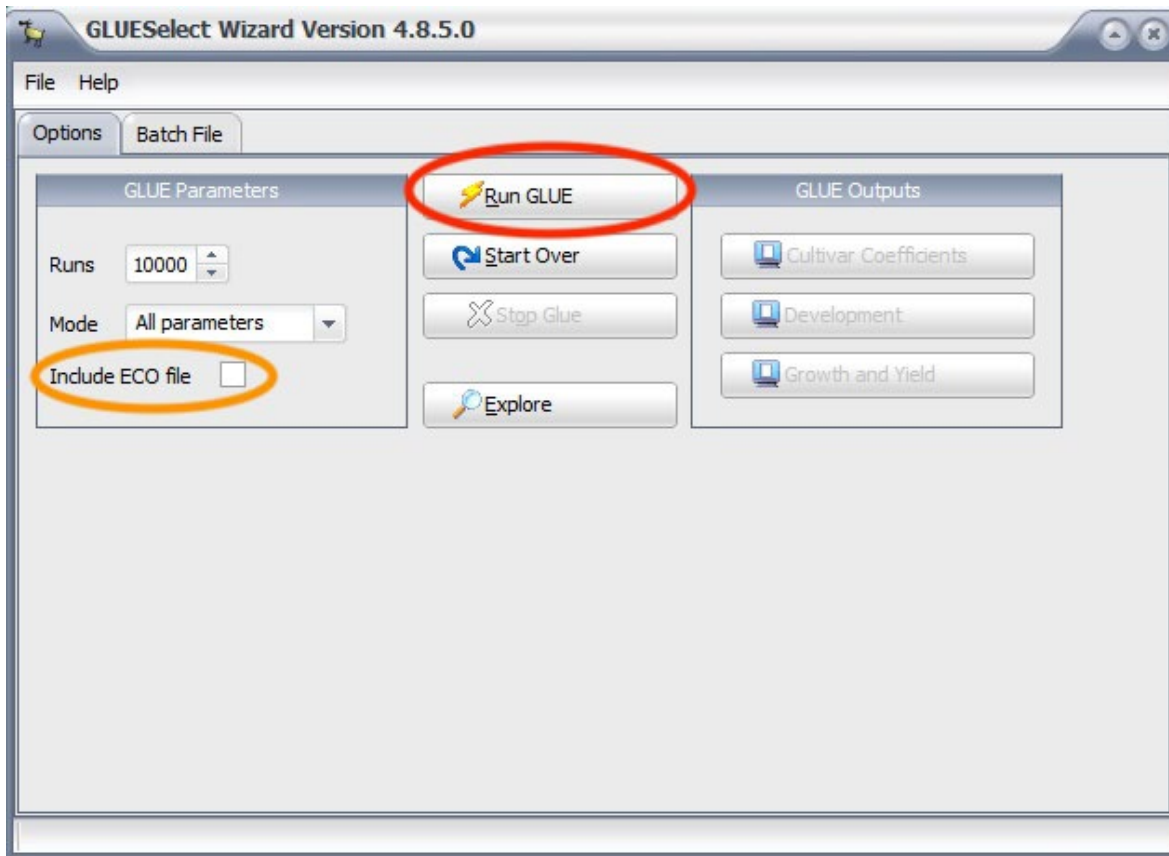
The second GlueSelect screen shows all available crops for calibration. A user selects a crop, such as peanut as shown at left, and then a particular cultivar that is to be estimated (“FLORUNNER, std” in this example). After selecting “Go” on this screen, a list of experiments and treatments will appear as shown below, as shown in the third GlueSelect screen below. In this example, one



The next GlueSelect screen shows the console for operating the GLUE calculations and viewing the results (below). In this example, 10,000 runs are set for all parameter groups, meaning 10,000 runs for phenology parameters and another 10,000 for growth parameters. The parameter groups can also be individual selected through the drop-down menu in ‘Mode’ This number can be changed to a few, say 10, to make sure that



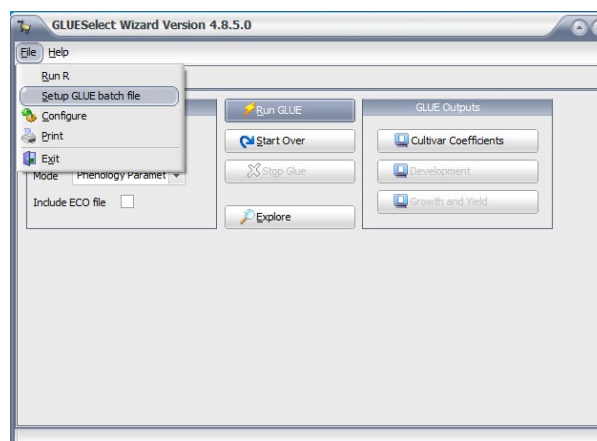
the program is operating ok. However, results from runs less than 10,000 would not likely give reliable results. So, if the number is changed from 10,000 to test the procedure, then it must be changed back to 10,000 (or more) in order to get reliable and stable results. Increasing the run count can further refine the results, though it will require more time. To include the calibration of genetic coefficients in the ecotype file, users must toggle the “Include ECO file” option, enabling GLUE to calibrate coefficients in both the cultivar and ecotype files. Note that not all crop growth models have defined parameter traits (P/G/N) or set ranges (MINIMA and MAXIMA) in the ecotype file. Therefore, make sure these definitions are present in the ECO file before selecting this option.



3.2. Running GLUE

The figure above illustrates how users initiate simulation runs to estimate the coefficients using the “Run GLUE” button. Depending on the number of model runs defined and seasons selected, GLUE runs may take some time to complete. From this screen, users can view the final estimated coefficients and copy them to the appropriate cultivar file (and ecotype file if the “Include ECO file” option is enabled) located in the DSSAT48\Genotype directory. Additionally, users can review summary statistics for the estimated coefficients, including the mean, maximum likelihood, and standard deviation.

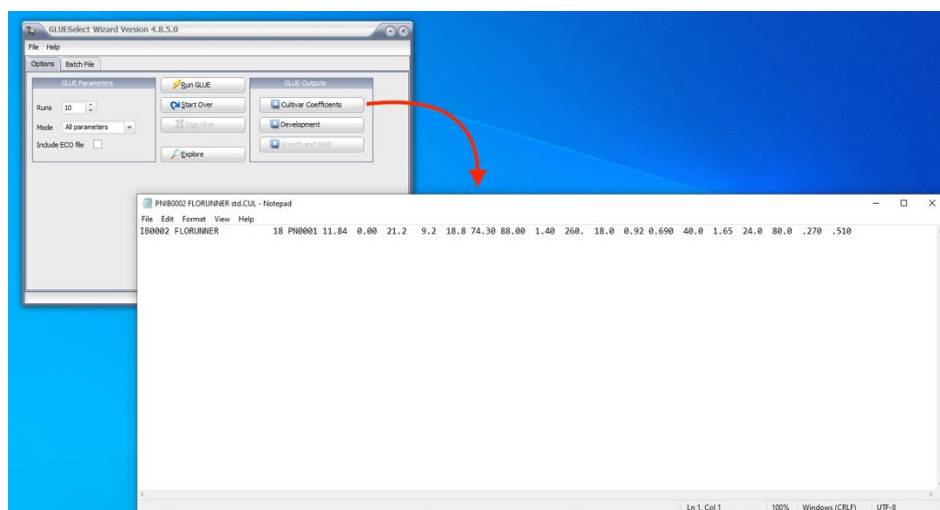
Occasionally, GlueSelect may be unable to automatically detect the correct path to the R program, which prevents it from running properly. When this occurs, you'll need to manually update the default path used by the system. To do this, go to the menu and select 'File' > 'Setup GLUE batch file', as illustrated in the figure on the left. This step ensures that GlueSelect can communicate with R correctly, allowing your analysis to proceed without issues.



Advanced model users can find detailed instructions for calibrating genetic coefficients in DSSAT-CSM crop growth models using GLUE via the terminal/command prompt at <https://github.com/DSSAT/GLUE>.

3.3 GLUE results and outputs

The main results that users will be interested in can be seen by selecting the "Cultivar Coefficients" button on the main screen, as shown in the figure below. 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 are saved in the “C:\DSSAT48\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 C).

(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 C) 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 “RandomParameterSetsAndProbability_2.txt” (Table 4 and 5, Appendix C) store the really used parameter sets and their corresponding probabilities 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 “DSSAT48\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 cultivar and ecotype files 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 “MeasurementVariance_All.csv” file. The “MeasurementVariance_All.csv” file shows how additional sheets in the spreadsheet must be set up for the crop.

References

- Ferreira, T. B., Shelia, V., Porter, C., Cadena, P. M., Cortasa, M. S., Khan, M. S., Pavan W., Hoogenboom, G., 2024. Enhancing crop model parameter estimation across computing environments: Utilizing the GLUE method and parallel computing for determining genetic coefficients. *Computers and Electronics in Agriculture*, 227, 109513.
- Hoogenboom, G., Porter, C. H., Boote, K. J., Shelia, V., Wilkens, P. W., Singh, U., White J.W., Asseng S., Lizaso J.I., Moreno L.P., Pavan W., Ogoshi R., Hunt L.A., Tsuji G.Y., Jones, J. W., 2019. *The DSSAT crop modeling ecosystem*. In *Advances in crop modelling for a sustainable agriculture* (pp. 173-216). Burleigh Dodds Science Publishing.
- He, J., Jones, J. W., Graham, W. D., Dukes, M. D., 2010. Influence of likelihood function choice for estimating crop model parameters using the generalized likelihood uncertainty estimation method. *Agricultural Systems*, 103(5), 256-264.
- Hunt, L. A., Pararajasingham, S., Jones, J. W., Hoogenboom, G., Imamura, D. T., Ogoshi, R. M., 1993. GENCALC: Software to facilitate the use of crop models for analyzing field experiments. *Agronomy Journal*, 85(5), 1090-1094.
- Jones, J. W., Hoogenboom, G., Porter, C. H., Boote, K. J., Batchelor, W. D., Hunt, L. A., Wilkens, P.W., Singh, U., Gijsman, A.J., Ritchie, J. T., 2003. The DSSAT cropping system model. *European Journal of Agronomy*, 18(3-4), 235-265.
- Memic, E., Graeff, S., Boote, K. J., Hensel, O., Hoogenboom, G., 2021. Cultivar coefficient estimator for the cropping system model based on time-series data: a case study for soybean. *Transactions of the ASABE*, 64(4), 1391-1402.

Appendix

Appendix A 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).
PDIT	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)
LSM	3	9		2	Leaf number per stem at maturity. The symbol "#" was changed to ".", since it is the symbol of comments in R.

Appendix B Batch File "FLORUNNER,_std.PNC" Created with GLUESelect

\$BATCH (CULTIVAR) : PNIB0002 FLORUNNER, std					
@FILEX	TRTNO	RP	SQ	OP	CO
C:\DSSAT48\Peanut\UFGA8101.PNX	2	0	0	0	0

Appendix C

Output Files of GLUE Procedure

1. Optimal parameter set saved as a CUL file (PNIB0002 FLORUNNER, std.CUL)

IB0002 FLORUNNER 18 PN0001 11.84 0.00 21.2 9.2 18.8 74.30 88.00 1.40 260. 18.0 0.92 0.690 40.0 1.65 24.0 80.0 .270 .510

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

Param	CSDL	PPSEN	EM.FL	FL.SH	FL.SD	SD.PM	FL.LF	LFMAX	SLAVR	SIZLF	XFRT	WTPSD	SFDUR	SDPDV	PODUR	THRSH	SDPRO	SDLIP
Mean	11.84	0	22.286	8.9	18.558	78.2	88	1.03	375	190	1	0.158	23	1.9	24	80	0.27	0.51
STDEV	0	0	3.602	1.4	2.587	10.91	0	0	0	0	0	0	0	0	0	0	0	0
MaxProbability	11.84	0.00	21.2	9.2	18.8	74.30	88	1.03	375	190	1	0.158	23	1.9	24	80	0.270	0.510

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

Param	CSDL	PPSEN	EM.FL	FL.SH	FL.SD	SD.PM	FL.LF	LFMAX	SLAVR	SIZLF	XFRT	WTPSD	SFDUR	SDPDV	PODUR	THRSH	SDPRO	SDLIP
Mean	11.84	0	22.286	8.9	18.558	78.2	88	1.2	264	17.3	0.93	0.732	38.2	1.63	24	80	0.27	0.51
STDEV	0	0	3.602	1.4	2.587	10.91	0	0.21	7.2	0.9	0.08	0.21	2.6	0.365	0	0	0	0
MaxProbability	11.84	0.00	21.2	9.2	18.8	74.30	88	1.40	260	18.0	0.92	0.690	40	1.65	24	80	0.270	0.510

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	0.000	23.831	8.900	20.113	52.635	88.000	1.030	375.000	13.690	1.000	0.384	17.165	1.900	24.000	80.000	0.270	0.510	0.227
	12.549	0.000	26.064	8.720	21.883	69.549	88.000	1.030	375.000	13.690	1.000	0.384	17.602	1.900	24.000	80.000	0.270	0.510	0.117
	11.791	0.000	26.731	9.942	15.802	54.909	88.000	1.030	375.000	13.690	1.000	0.384	17.317	1.900	24.000	80.000	0.270	0.510	0.115
	11.895	0.000	27.931	8.781	16.092	74.792	88.000	1.030	375.000	13.690	1.000	0.384	17.585	1.900	24.000	80.000	0.270	0.510	0.103
	11.784	0.000	21.206	7.886	13.915	59.413	88.000	1.030	375.000	13.690	1.000	0.384	18.123	1.900	24.000	80.000	0.270	0.510	0.095
	12.339	0.000	19.642	5.194	19.395	78.400	88.000	1.030	375.000	13.690	1.000	0.384	17.539	1.900	24.000	80.000	0.270	0.510	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	0.000	23.831	8.900	20.113	52.635	88.000	1.090	246.041	13.690	1.000	0.384	17.165	1.900	24.000	80.000	0.270	0.510	0.016
	12.538	0.000	23.831	8.900	20.113	52.635	88.000	1.029	242.632	19.759	0.956	0.386	17.602	1.775	24.000	80.000	0.270	0.510	0.015
	12.538	0.000	23.831	8.900	20.113	52.635	88.000	1.110	230.573	14.342	0.678	0.386	17.317	1.564	24.000	80.000	0.270	0.510	0.015
	12.538	0.000	23.831	8.900	20.113	52.635	88.000	1.011	277.951	13.991	0.724	0.390	17.585	2.000	24.000	80.000	0.270	0.510	0.015
	12.538	0.000	23.831	8.900	20.113	52.635	88.000	1.082	259.885	15.893	0.778	0.394	18.123	1.792	24.000	80.000	0.270	0.510	0.015
	12.538	0.000	23.831	8.900	20.113	52.635	88.000	1.150	275.298	13.908	0.891	0.391	17.539	1.719	24.000	80.000	0.270	0.510	0.014

