User guidelines for running TSE
(Time-Series Estimator – for DSSAT models) software

Emir Memic[1]

[1] University of Hohenheim, Stuttgart, Germany, August 2020

[1]E-mail: emir.memic@uni-hohenheim.de; emir_memic@windowslive.com

This cultivar coefficients optimisation program was developed and tested with DSSAT4.7 CROPGRO-Soybean model.

The program was partially tested with CERES-Maize and Wheat models.

The program was designed and written to enable optimisation of cultivar coefficients of all available crop models in DSSAT4.7 shell.

Contents

1. The TSE concept overview

The cultivar coefficients estimation program was developed to work with DSSAT crop growth models (CROPGRO and CERES). It will work only if fully functional DSSAT files are available such as experiment file (FileX), weather file, soil profile, functional species, ecotype and cultivar files, time-series file containing in-season observations (FileT) and summary file (FileA - optional) are available.

Overall program run can be separated in three steps:

1. Selection of the FileX and corresponding treatments defined in FileX. Based on the selected experiment corresponding in-season observations (as time-series data) are red in as temporal inputs (all available time-series observations can be used), and cultivar coefficients designated for optimisation (phenology- and/or growth-related) are selected with desired coefficient ranges (Figure 1, step 1).

2. For each cultivar coefficient combination defined in first step crop model is executed with simulated outputs being coupled with in-season observations and saved for later analysis (Figure 1, step 2).

3. After all cultivar coefficient combinations were used and simulated outputs coupled with in-season observations statistical analysis is conducted based on the target variables used in the cultivar coefficient estimation process and the combination with the best statistical agreement between simulated and observed is selected as "optimum".

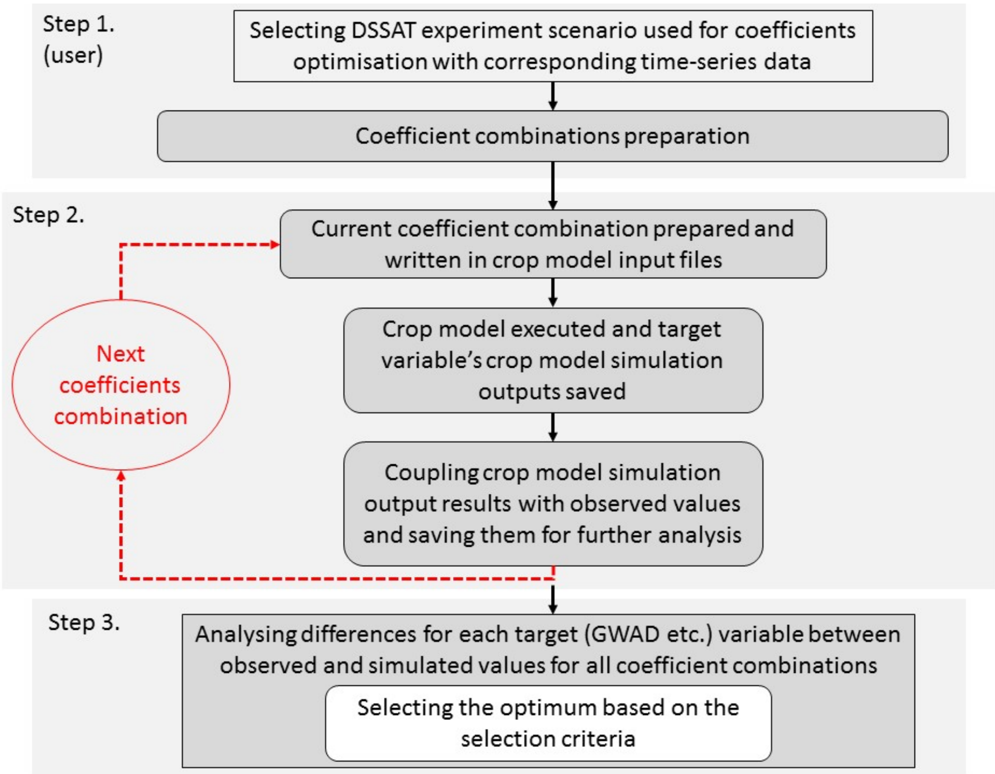


Figure 1: Flow chart showing overall program run in three steps (GWAD – grain weight) (Memic et al. 2020)

The program can be used for optimising cultivar coefficients based on single or multiple experiment data sets. User should first optimise phenology-related cultivar coefficients. After phenological events are correctly optimised, growth-related cultivar coefficients are optimised. Phenology-related cultivar coefficient optimisation is not conducted based on the time-series data (FileT), but by using FileA observations and simply minimising the difference between simulated and observed phenological event as day after planting. Growth-related cultivar coefficients are optimised by using FileT in-season observations by means of normalised RMSE throughout season. The combination providing lowest difference between simulated and observed values based on nRMSE is selected as optimum. Cultivar coefficient ranges initial values are predefined (by existing minimum and maximum values in cultivar files) but can be modified according to user's needs.

2. General TSE program settings overview

The program was developed (in Windows environment, for use on Windows) as standalone and has no specific installation requirements. All required files for running the program are shared with gitHub_TSE.zip file.

Before using the program user should look into User Guidelines shared with the program.

The "gitHub_TSE_[date_stamp].zip" file has to be unzipped as "gitHub_TSE_[date_stamp]" working directory. From "gitHub_TSE_[date_stamp]" directory "**TSE**" folder copied to the Tools directory: "C:\DSSAT47\Tools" (Figure 2).

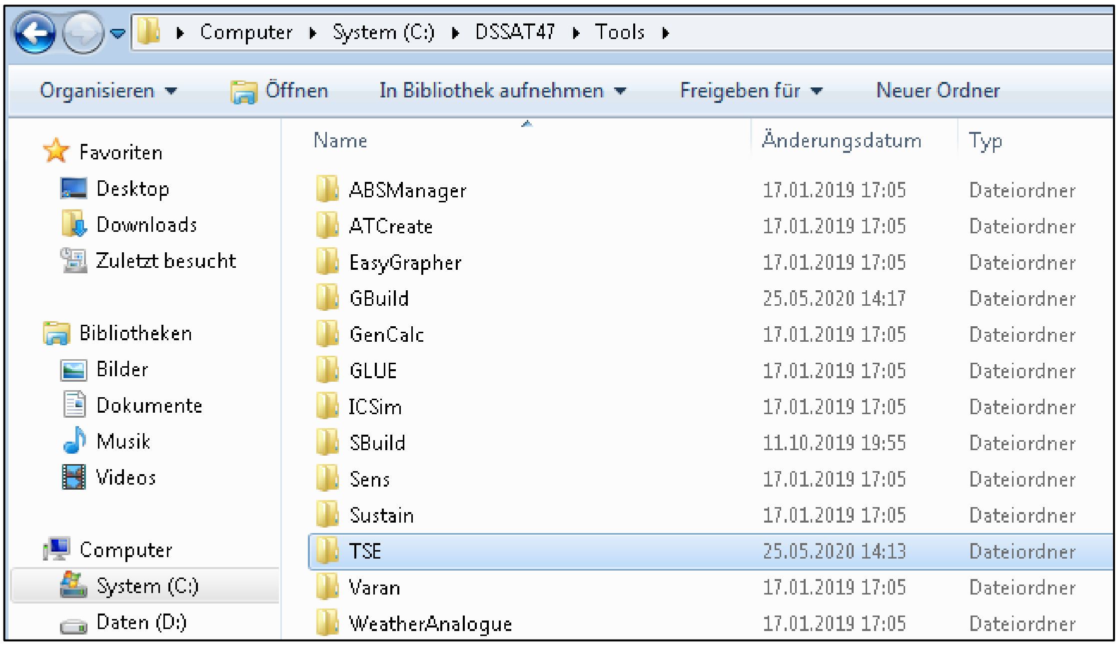


Figure 2 "C:\DSSAT47\Tools\TSE"

In folder "TSE" "C:\DSSAT47\Tools\TSE" (Figure 3) **TSE_calibrator_DSSAT.exe** windows runnable has to be **executed as "Administrator"** (Figure 4).

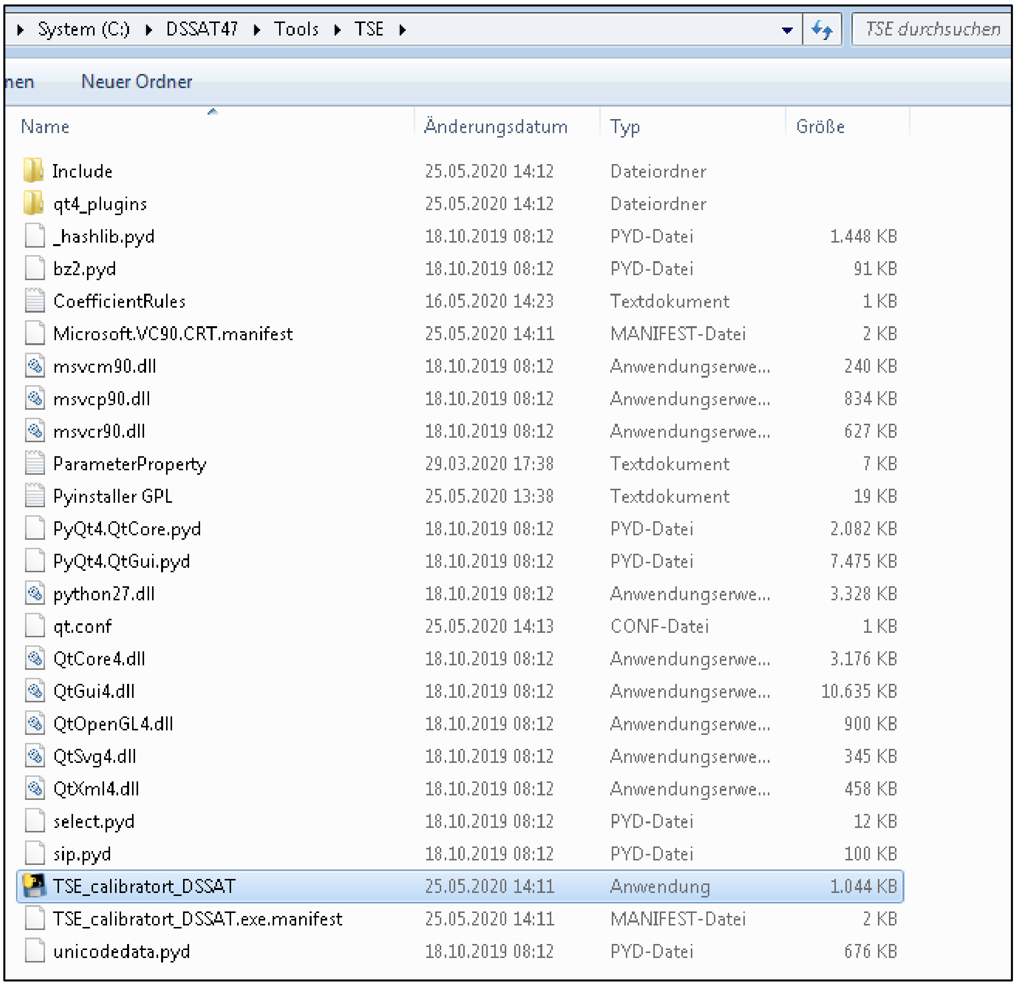


Figure 3 **TSE_calibrator_DSSAT.exe** windows runnable

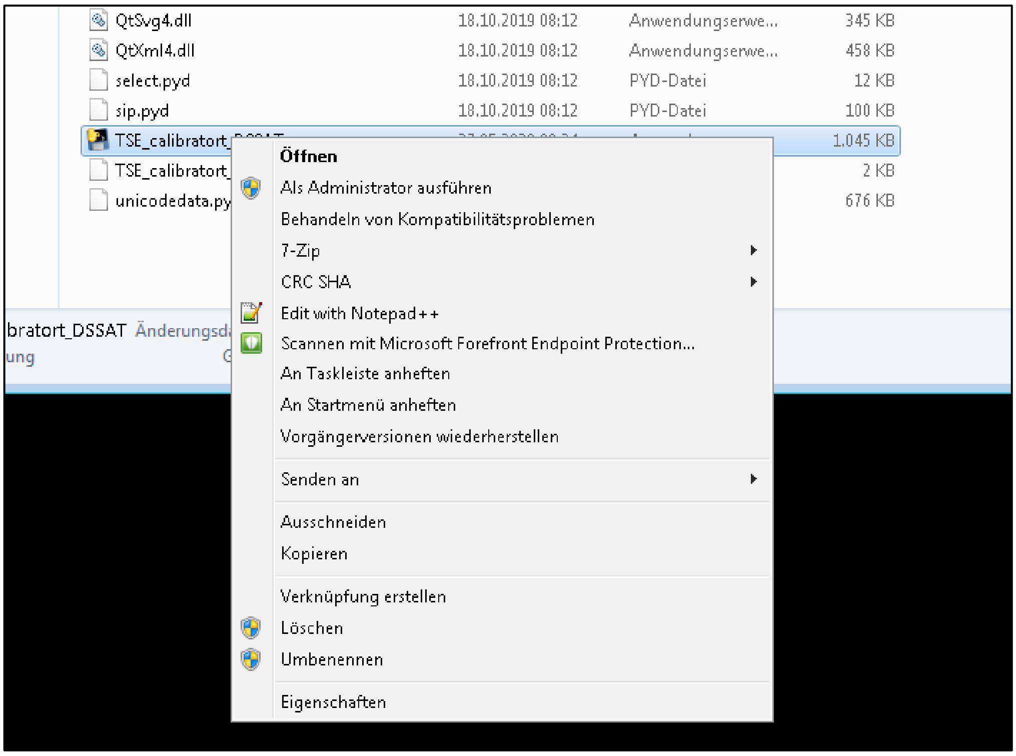


Figure 4 **Execute as “<u>Administrator</u>”**

VERY IMPORTANT:

I. PlantGro.Out crop model outputs are coupled to those in FileT (growth-related) time-series in-season observations
II. Evaluate.Out crop model outputs are coupled to those in FileA (phenology-related) as DAY observations
III. If sub-model (eg. WHAPS) is initialised in the FileX, the calibrator will not work! (in FileX in *SIMULATION CONTROLS in GENERAL line, column SMODEL **do NOT initialise sub-models!**)
IV. Only variables such as LAID, CWAD, GWAD etc. initialised in first time occurring "**@TRN….**" line in FileT is actively used by TSE
V. For multi TRT optimisations only target variables simultaneously available in all FileT/s (for corresponding FileX/s Treatment/s) are accessible for optimisation
VI. If PrameterProperty.txt is NOT in the C:\DSSAT47\Tools\TSE folder, **P**/**G** (**P**henlology/**G**rowth-related flags) are not available in the interface!
VII. Tfile observations (all in-season observations available including 0 are used, only -99 values are ignored by the program) used for estimating the optimum genetic coefficient (phenology- and growth-related)
VIII. The program is matching DOY from FileT with those in the PlantGro.OUT. If the user setup in the FileX reporting frequency for example every fifth day and exact observation DOY is not present in the PlantGro.OUT as it is written in the FileT, the program will not be able to match them for comparing simulated with observed.

The optimiser program is creating additional directory "**TSE_workspace**" (**C:\DSSAT47\TSE_workspace**) (Figure 5) and modifying the cultivar file in that directory, which is then executed by main DSSAT model executable.

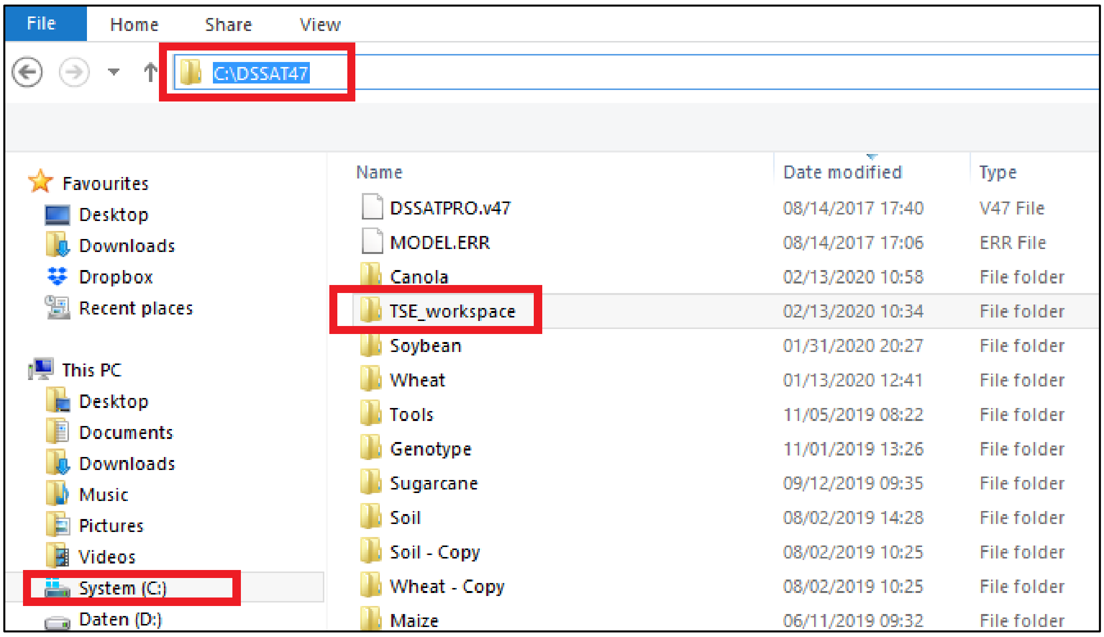


Figure 5 **C:\DSSAT47\TSE_workspace**

The optimiser program is NOT modifying core DSSAT files in their original directories!

Program run is considered: Start- "TSE_calibrator_DSSAT.exe" executed until "Exit" push button is pressed. Any form of optimisation done in-between is temporary saved in the temporary cultivar file in TSE_workspace directory.

After TSE program is started (**TSE_calibrator_DSSAT.exe executed**) all modifications on Cultivar file are conducted in **C:\DSSAT47\TSE_workspace**. During one program run (until "**Exit**" push button is pressed) different coefficients (or different target variables) can be optimised one after another or simultaneously and cultivar changes will be saved if accepted as "optimums" in cultivar file in **C:\DSSAT47\TSE_workspace**. If user is satisfied with the cultivar coefficient values based on nRMSE fit cultivar coefficient combination should be copied to **C:\DSSAT47\Tools\Genotype** located cultivar file, **MANUALLY**. If TSE program is started again without saving the combination in **C:\DSSAT47\Tools\Genotype** located cultivar file new TSE program start will copy original **C:\DSSAT47\Tools\Genotype** located cultivar file and overwrite user's working cultivar file in **C:\DSSAT47\TSE_workspace**.

After model run finished and before the user click "**Exit**" push button they can open GBuild and check visual and statistical fit (RMSE, d-statistics within GBuild) of the experiment file executed with the "optimum" genetic coefficient combination found in the last model run. With GBuild the user opens PlantGro.OUT from C:\DSSAT47\TSE_workspace directory, because TSE will create parallel files it requires in this folder, without modifying the original files in DSSAT directory.

The more coefficients are "activated" (used in estimation process) the longer will optimisation last. **For each new coefficient and additional increment step (Inc) number of model runs will increase exponentially**.

For example:

Every time TSE_calibrator_DSSAT.exe is executed, original cultivar (SBGRO047.CUL) file from C:\DSSAT47\Genotype will be copied to C:\DSSAT47\TSE_workspace directory (Figure 6), and overwrite cultivar file in that directory (if exist, if not then just copied). If user wants to keep the genetic coefficient combination, it has to be copied to the original cultivar file in C:\DSSAT47\Genotype directory into SBGRO047.CUL manually.

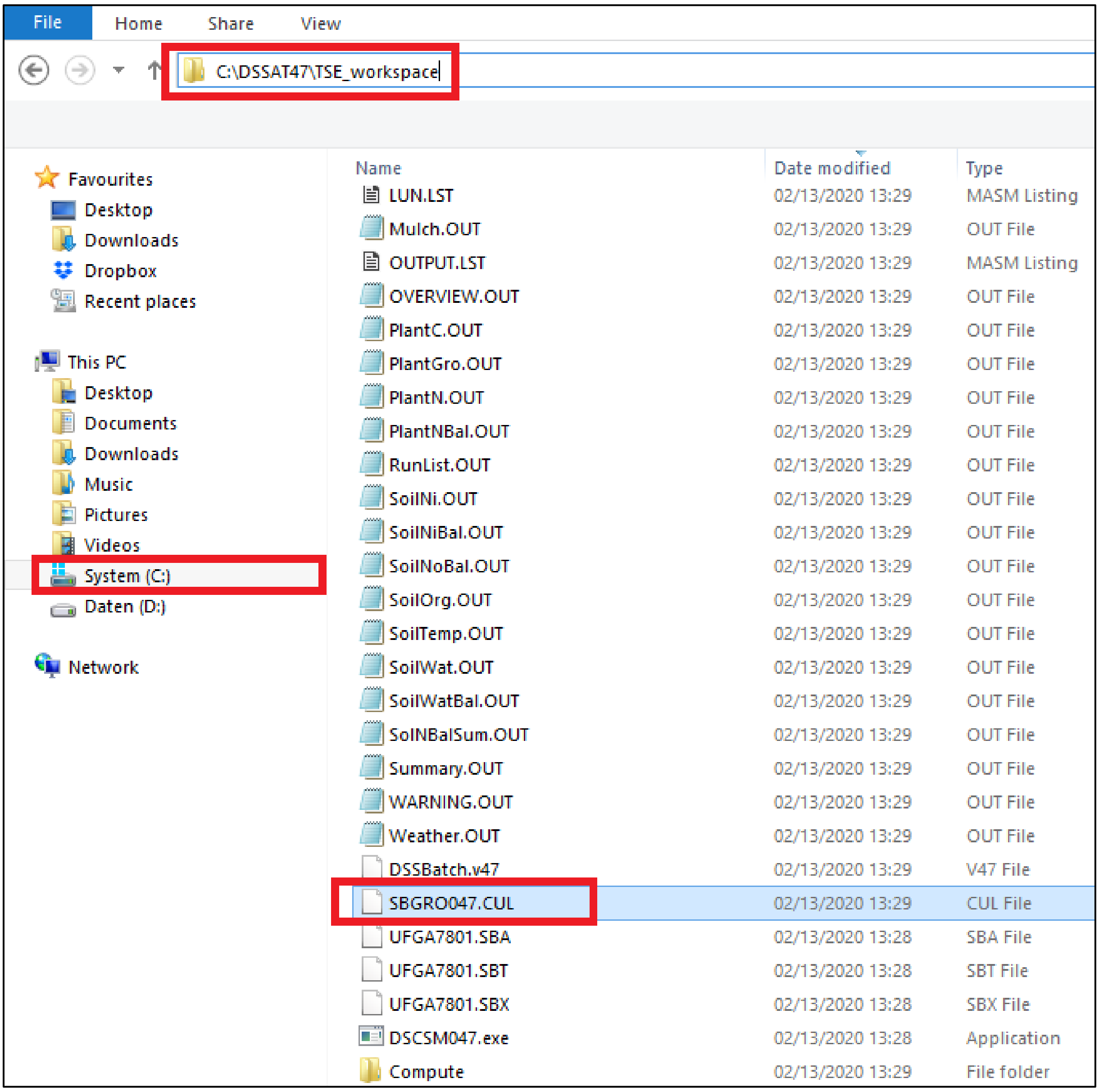


Figure 6 C:\DSSAT47\TSE_workspace\SBGRO047.CUL

The original cultivar is saved as **“!Old_timestamp_cultivarID….”** (Figure 7, text editor line 71) and the new one (Figure 7, text editor line 74) is saved in that working cultivar file (in **TSE_workspace**) and model is executed.

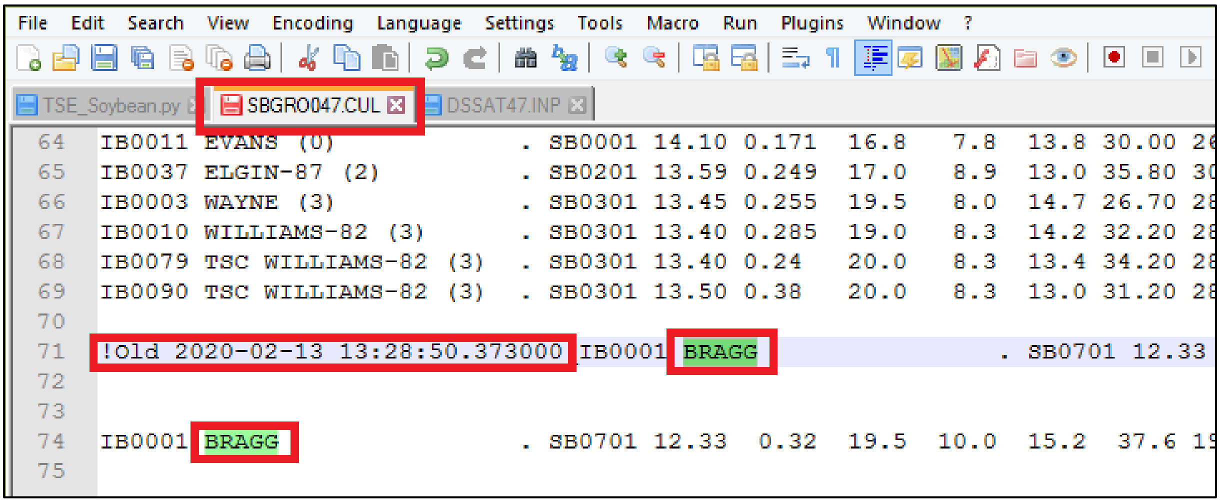


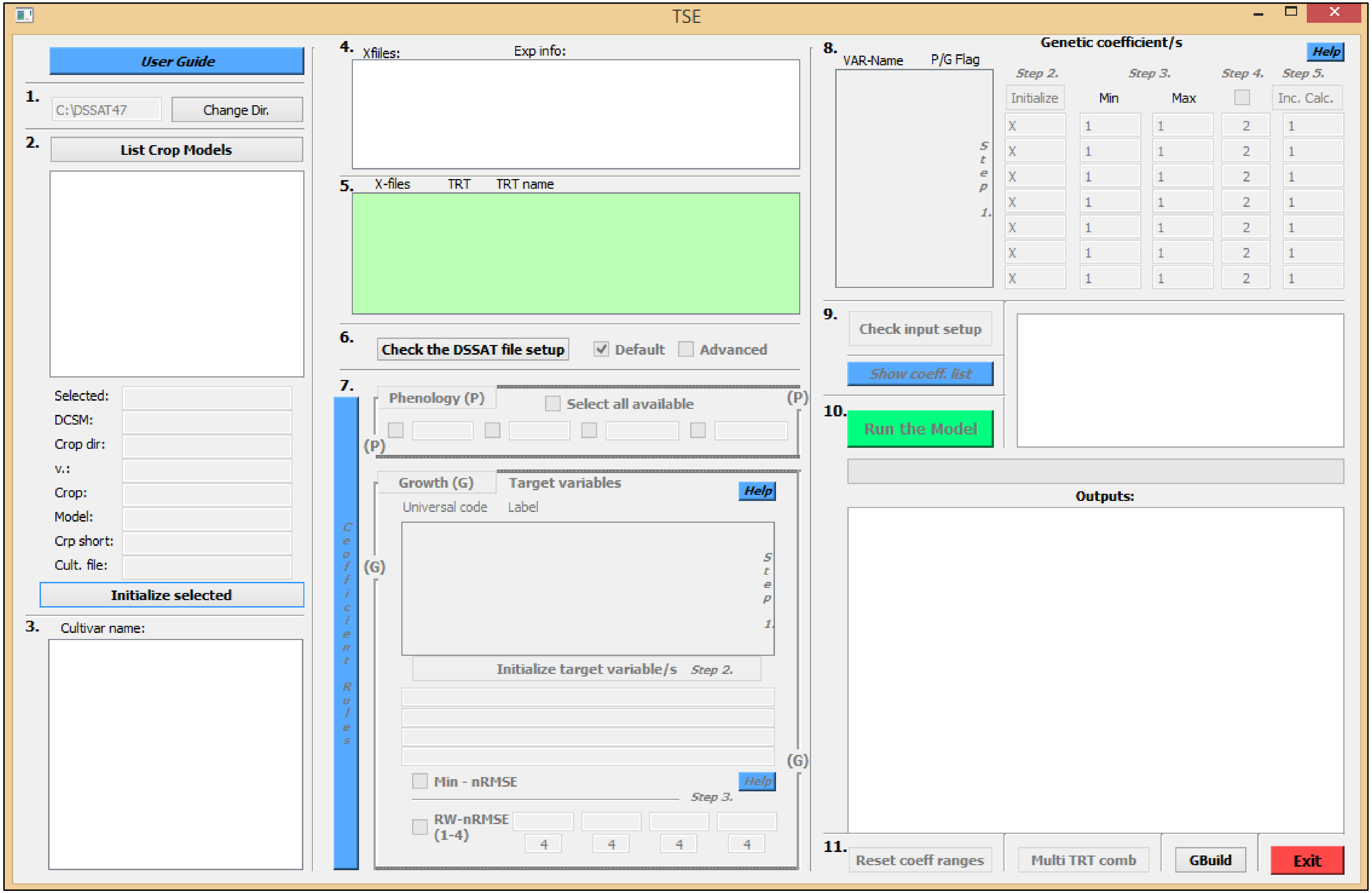
Figure 7 Old and new cultivar

3. Running TSE program

(The steps of preparing the estimator for run are enumerated in the interface)

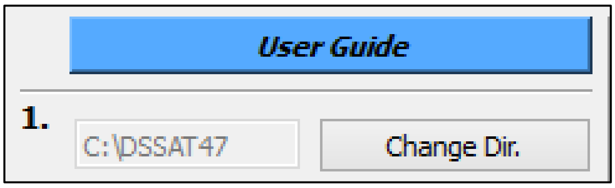
1. **If directory path shown is "C:\DSSAT47", do NOT modify! If the path is not "C:\DSSAT47" (This means that TSE folder was not copied to the "C:\DSSAT47\Tools"), then navigate to TSE folder and select it. It will be explained later in more details.**
2. **Select desired model and Initialize it!**
3. **Select cultivar from model corresponding list.**
4. **Select File/s-X from list containing selected cultivar.**
5. **Select corresponding Treatment/s based on the File/s-X containing selected cultivar.**
6. **Execute selected treatment/s with DSSAT model to check if core DSSAT files are runnable, and select Default/Advanced.**
7. **Select optimisation of Phenology/Growth -related coefficients and corresponding methods.**
8. **Selecting desired coefficients and coefficient ranges and increment steps.**
9. **Check if optimisation software setup is correct.**
10. **Run the model!**
11. **Reset coefficient ranges or estimate Multi treatment based cultivar coefficient combination!**

Interface


TSE
User Guide
1.
C:\DSSAT47
Change Dir.
2.
List Crop Models
Selected:
DCSM:
Crop dir:
v.:
Crop:
Model:
Crp short:
Cult. file:
Initialize selected
3.
Cultivar name:
4.
Xfiles:
Exp info:
5.
X-files
TRT
TRT name
6.
Check the DSSAT file setup
Default
Advanced
7.
Coefficient Rules
Phenology (P)
Select all available
(P)
(P)
Growth (G)
Target variables
Help
Universal code
Label
Step 1.
(G)
Initialize target variable/s
Step 2.
Min - nRMSE
Help
Step 3.
RW-nRMSE (1-4)
4
4
4
4
(G)
8.
VAR-Name
P/G Flag
Step 1.
Genetic coefficient/s
Help
Step 2.
Step 3.
Step 4.
Step 5.
Initialize
Min
Max
Inc. Calc.
9.
Check input setup
Show coeff. list
10.
Run the Model
Outputs:
11.
Reset coeff ranges
Multi TRT comb
GBuild
Exit

1. *Do NOT modify!*

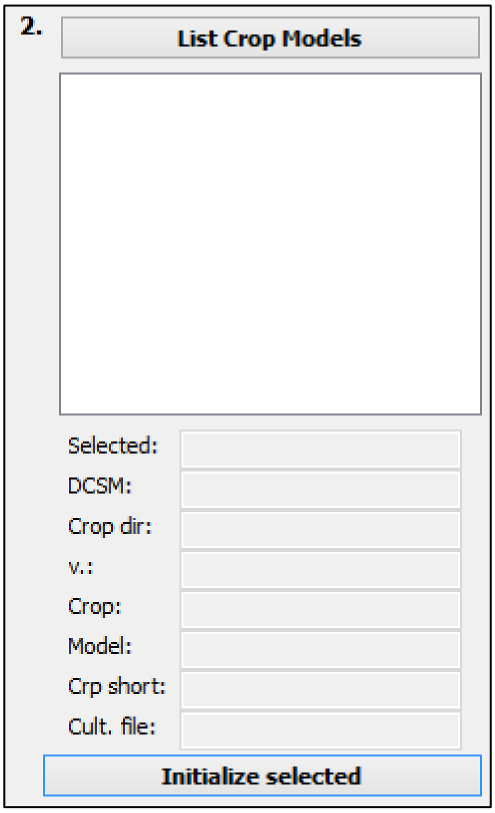
If directory path shown is "C:\DSSAT47", do ***NOT*** modify! If the path is not "C:\DSSAT47" then copy TSE folder (not entire gitHub_[date_satemp] extracted folder!, only TSE folder from it into the "C:\DSSAT47\Tools". After re-executing **TSE_calibrator_DSSAT.exe** from TSE **->** "C:\DSSAT47\Tools\TSE" the correct path should be red in.



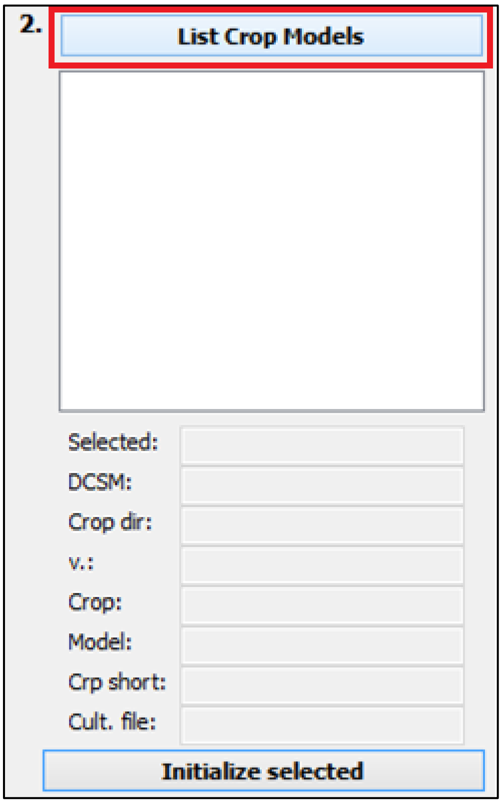
1.1 Interface

2. Select desired model and Initialize selected!

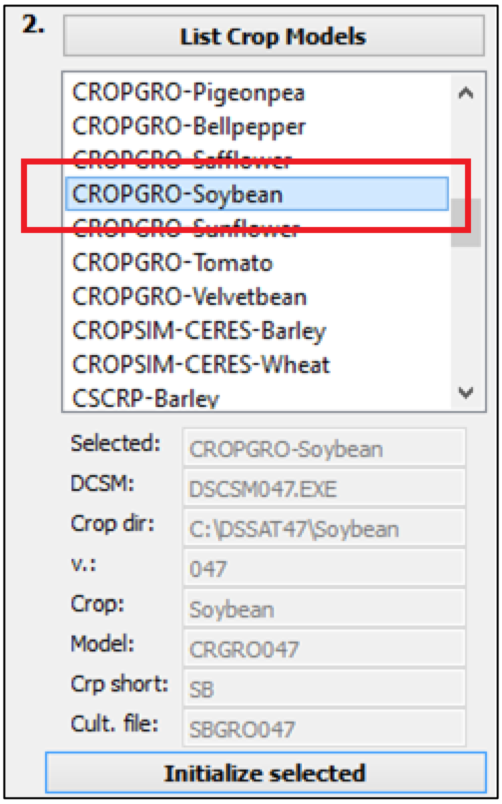
The model list is uploaded from “C:\DSSAT47” SIMULATION.CDE file.



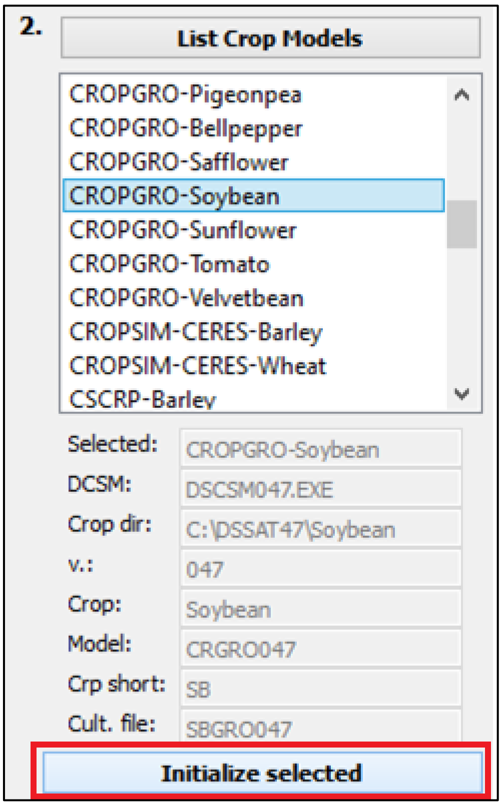
2.1 Interface



2.2 Interface

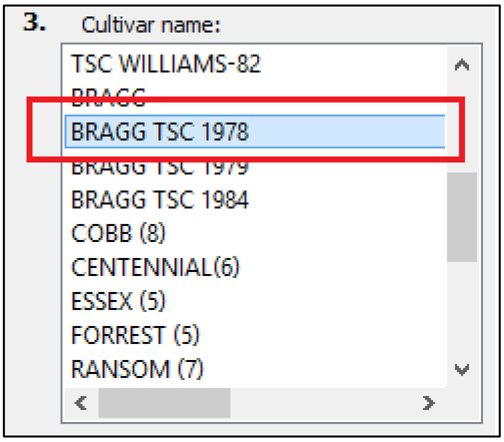


2.3 Interface

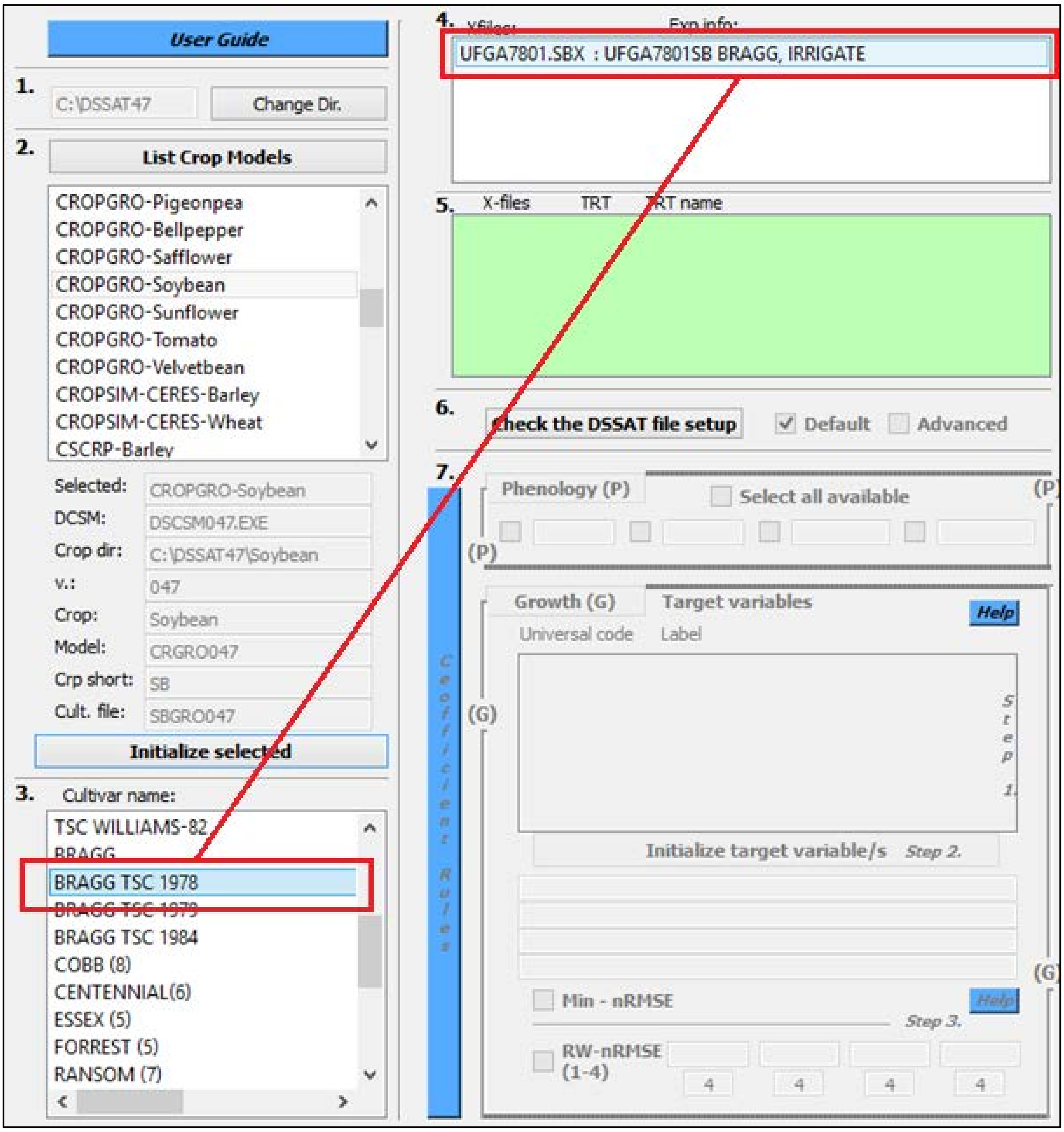


2.4 Interface

3. Select cultivar from model corresponding list.

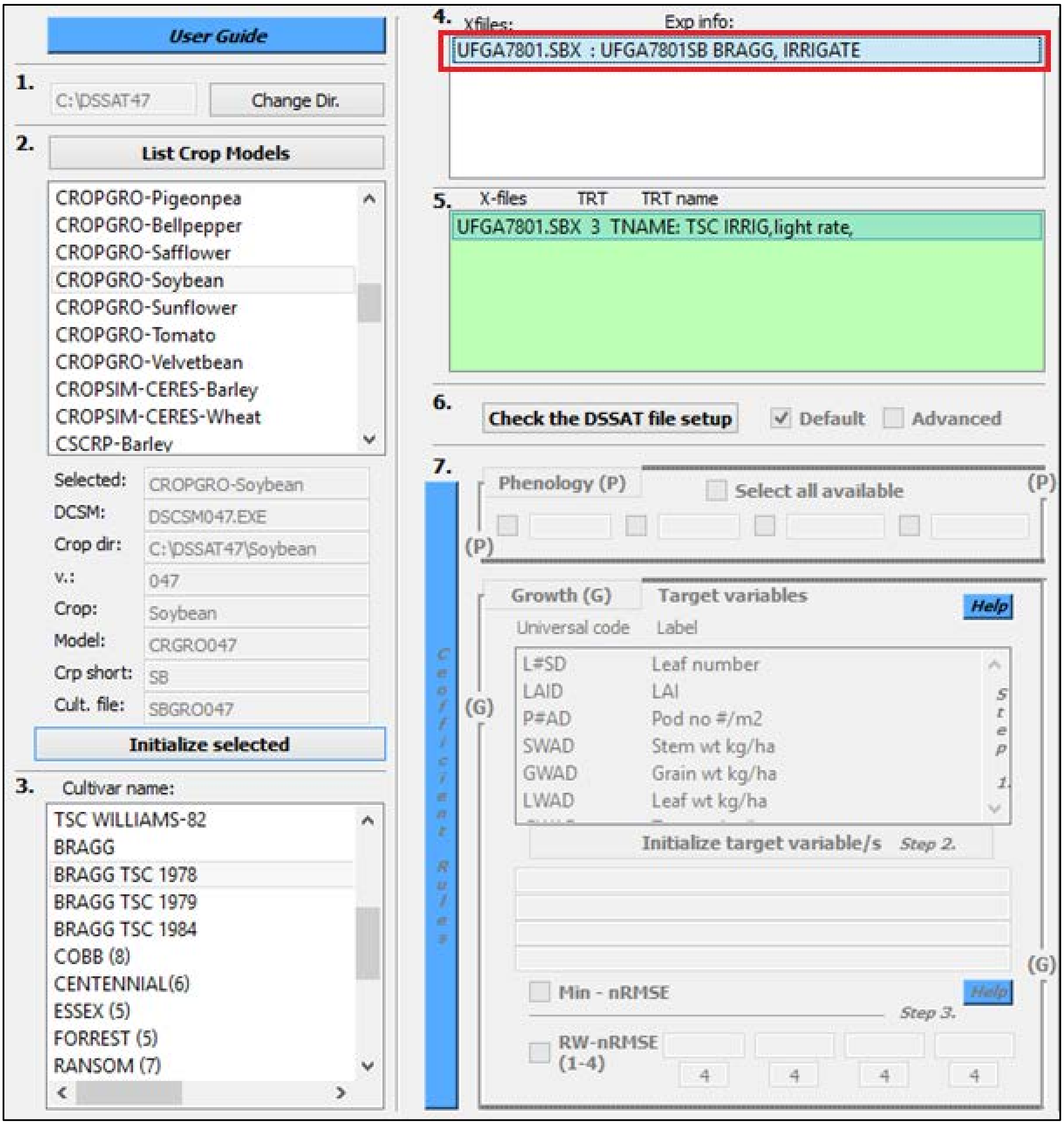


3.1 Interface



3.2 Interface

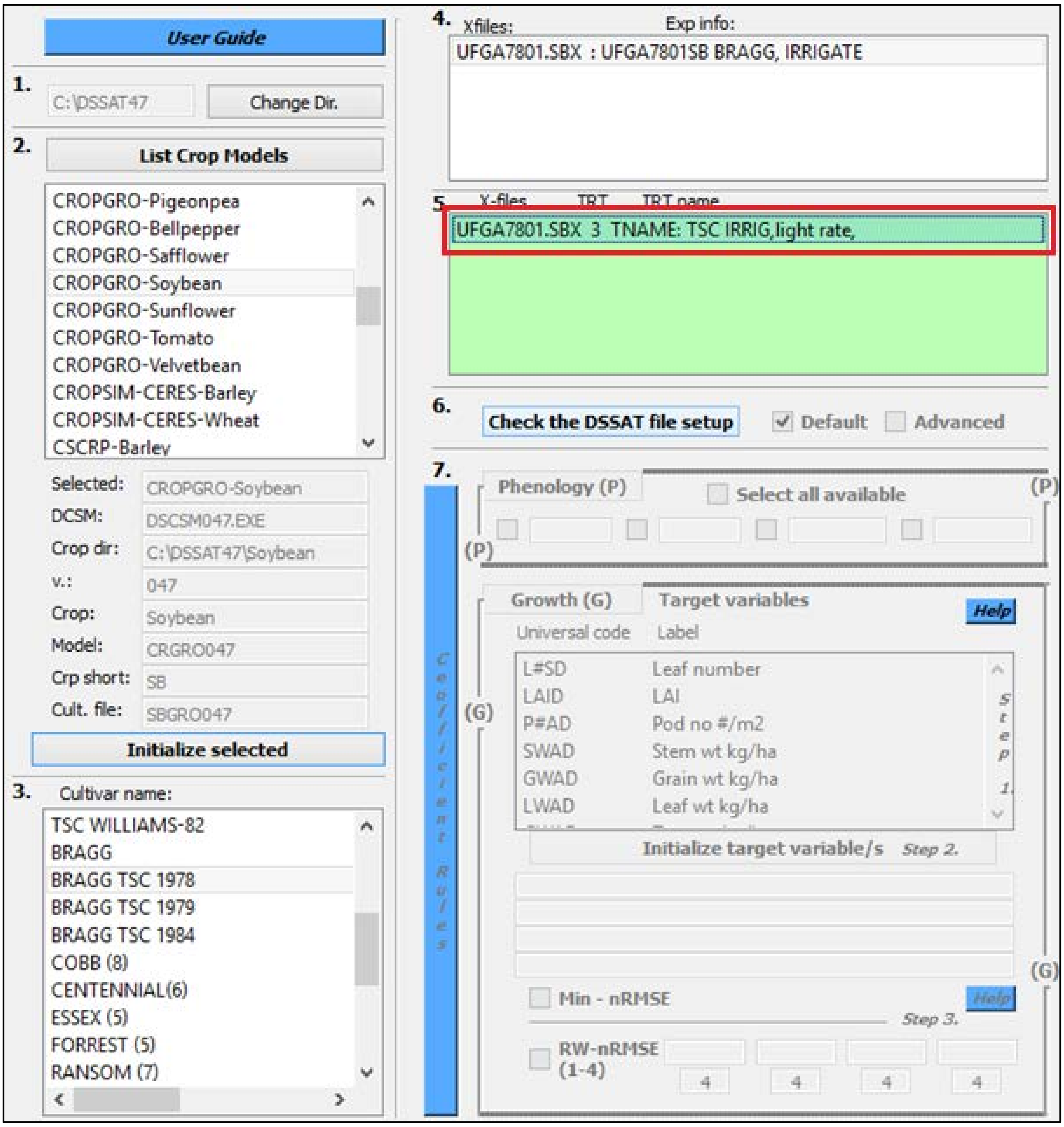
4. *Select File/s-X* *from list containing selected cultivar.*



4.1 Interface

5. Select corresponding Treatment/s based on the File/s-X containing selected cultivar.

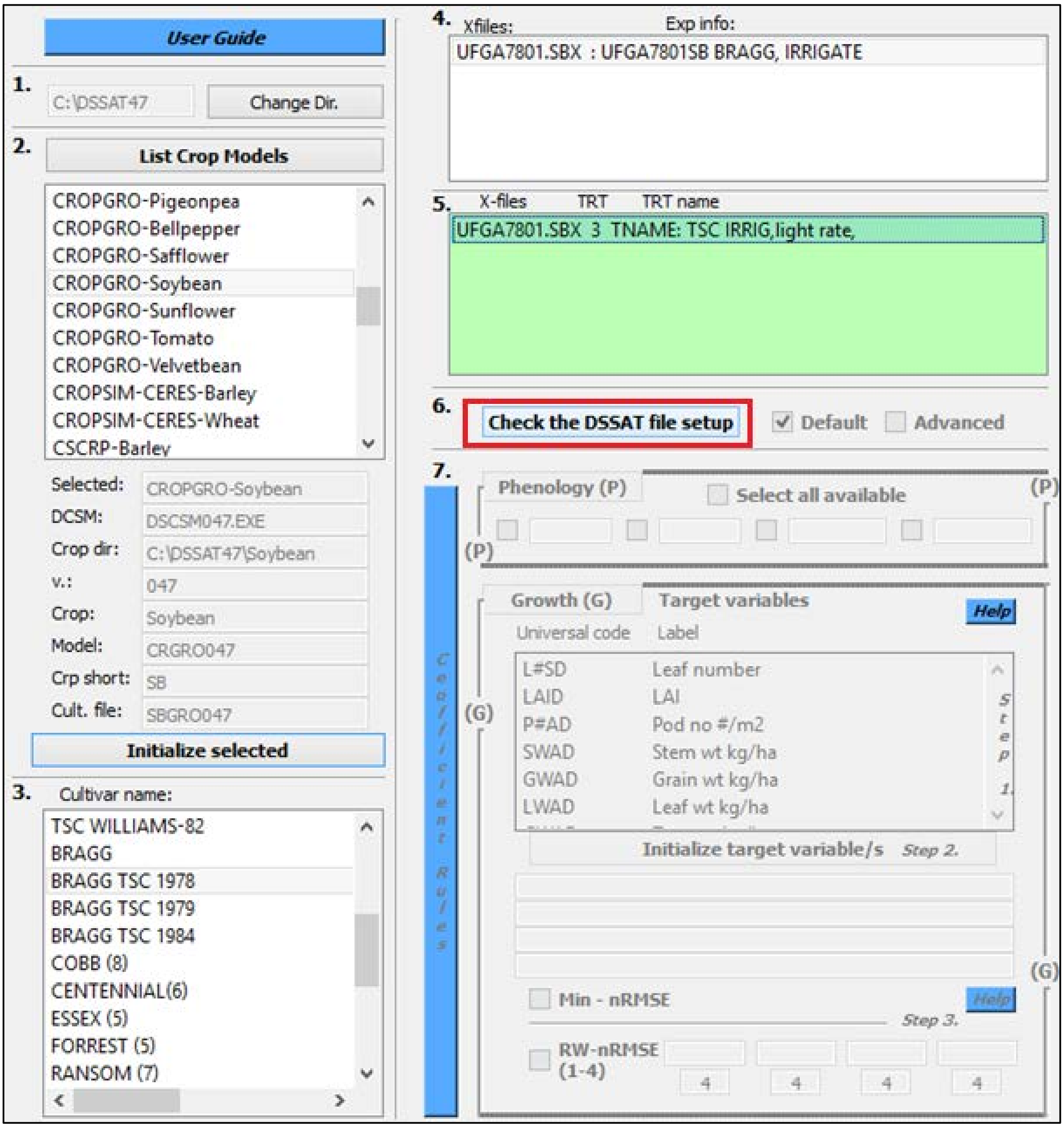
The green box (green list widget box in the interface) is where the experiment file and treatments for optimising cultivar coefficients are **selected**. Multiple treatments can be selected (multiple treatment/experiment selection is done by: **Ctrl+ mouse left button click)**.



5.1 Interface

***6.** Execute (Check the DSSAT file setup) selected treatment/s with DSSAT model to check if core DSSAT files are runnable, and select Default/Advanced.*

6.1 Interface



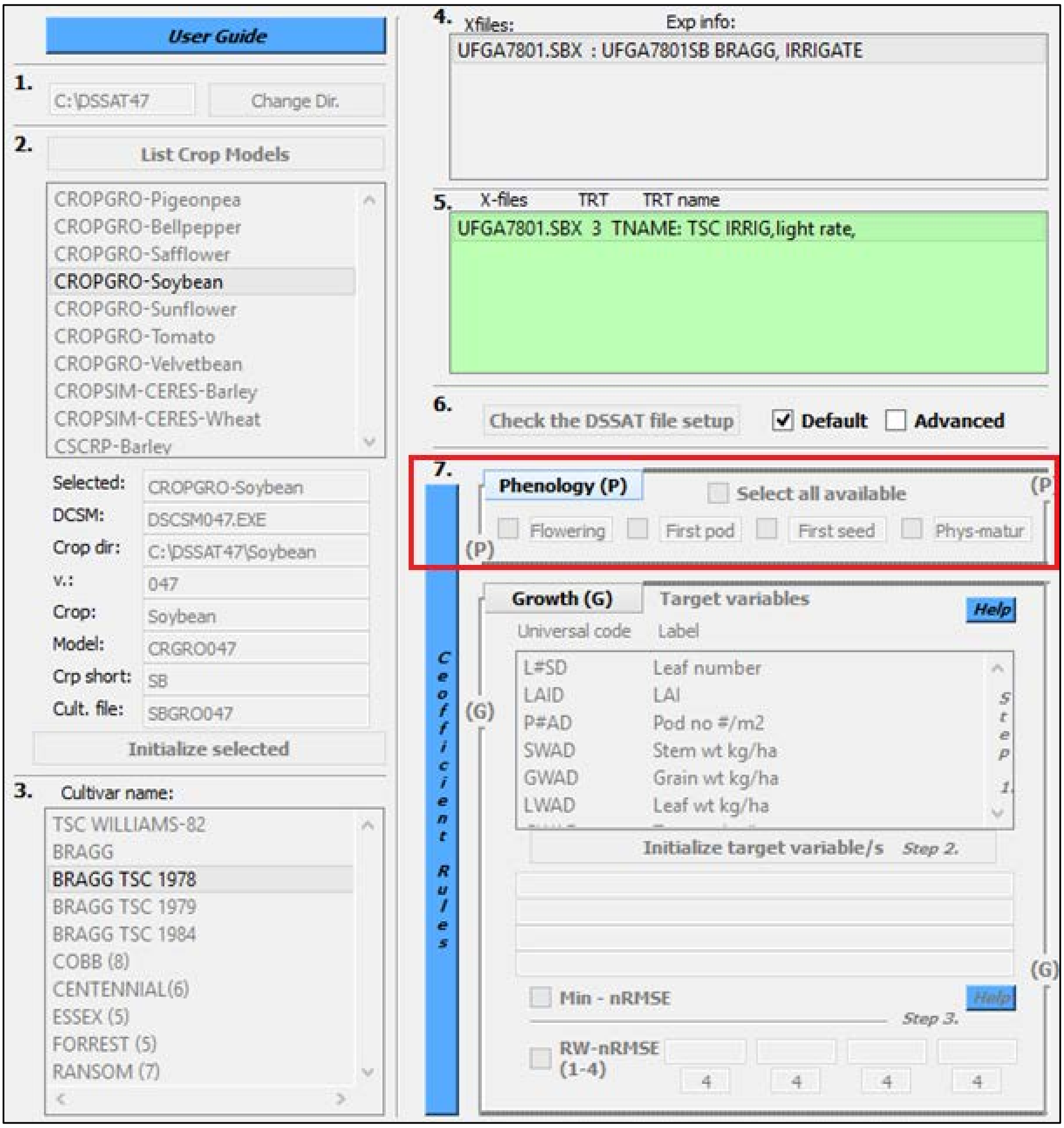
6.2 Interface

Cultivar specific coefficient flags (Phenology and Growth – P/G) are listed in the "C:\DSSAT47\Tools\TSE" in a file "ParameterProperty.txt". ***Default*** (check box) will upload only coefficients with predefined P/G flags (according to selection of the Phenology or Growth). ***Advanced*** will upload all available cultivar coefficients from cultivar file for potential optimisation.

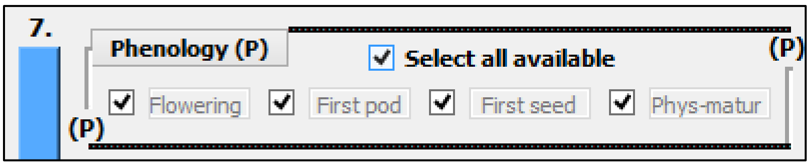
7. Select optimisation of Phenology/Growth -related coefficients and corresponding methods.

In this step error minimisation method (nRMSE or RW-nRMSE, explained in the APPENDIX) is selected. Maximum four different target variables (LWAD, CWAD, PWAD and SHAD) can be selected. If four are not available, one or two or three can be selected as well. In first round of optimisation four can be selected and in second next four, or with some other combinations.

Phenology:



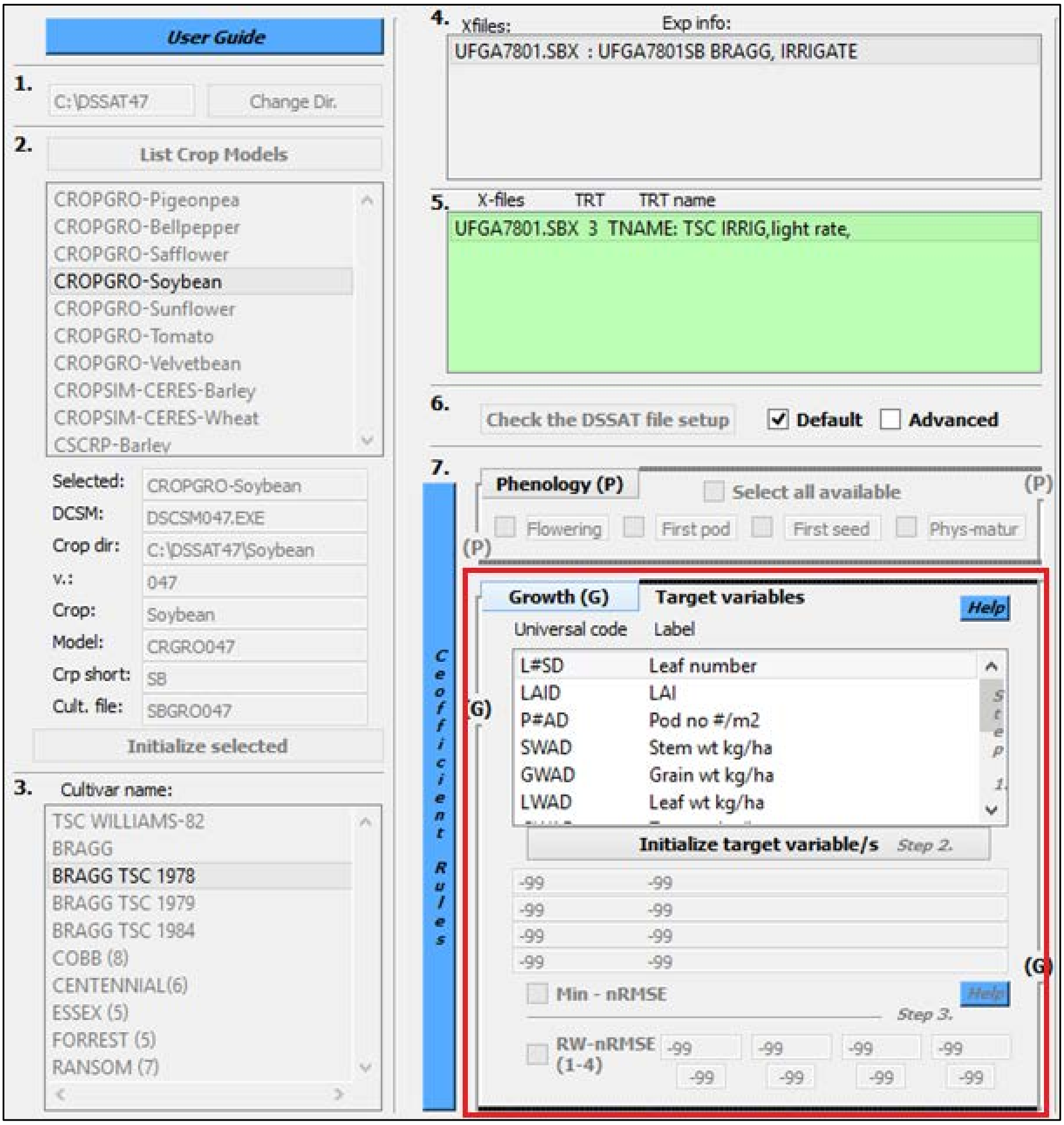
7.1 Interface



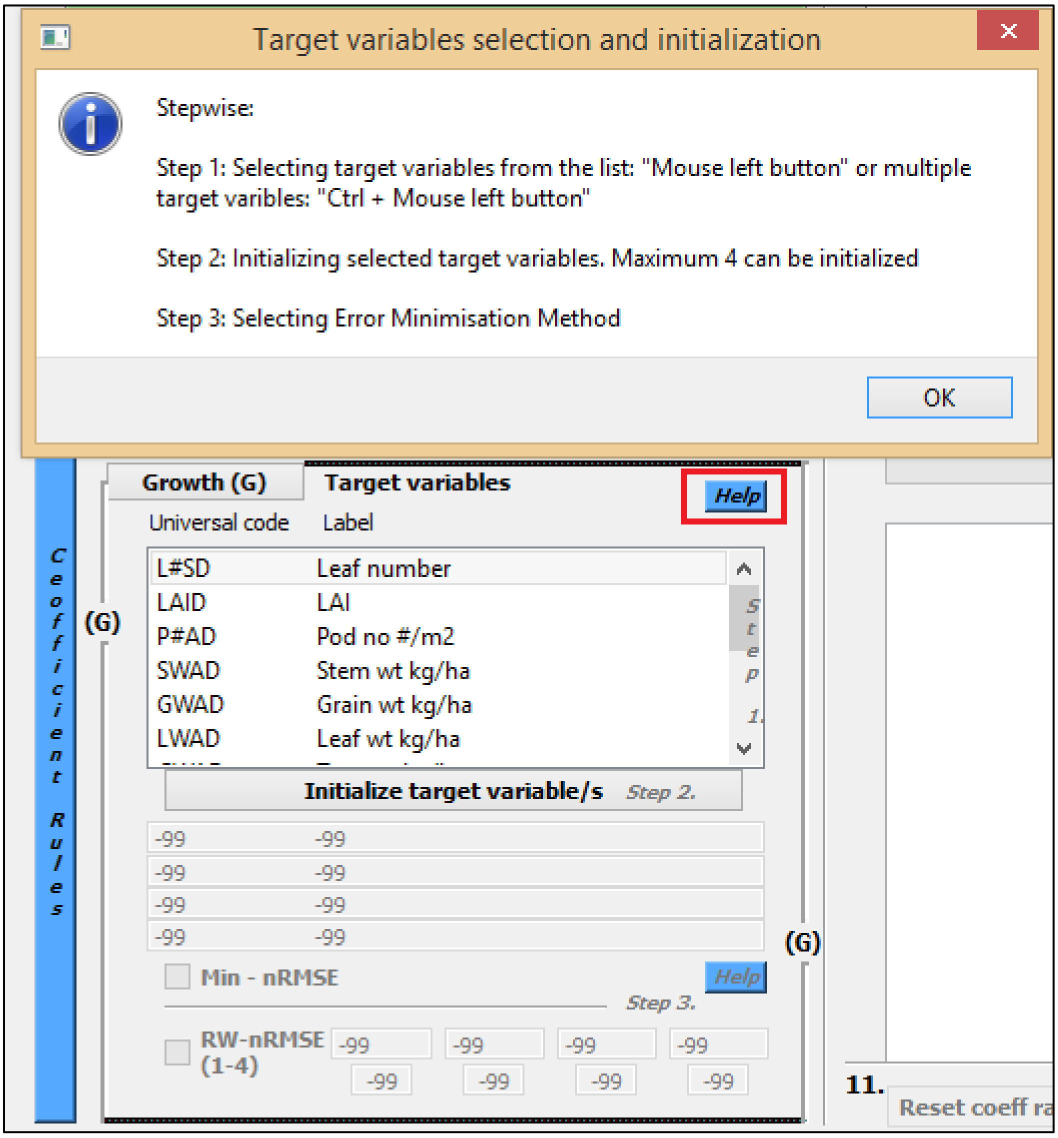
7.2 Interface

7.3 Interface

Growth:

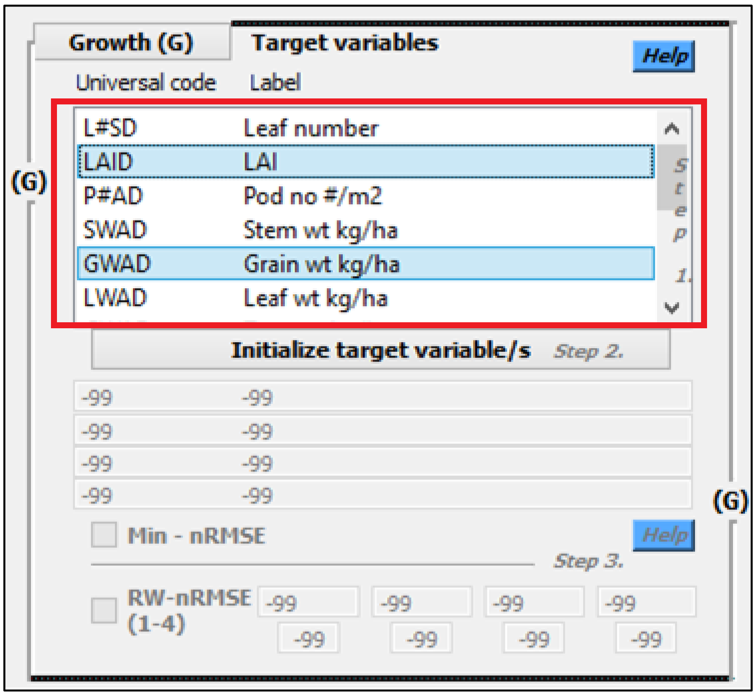


7.4 Interface

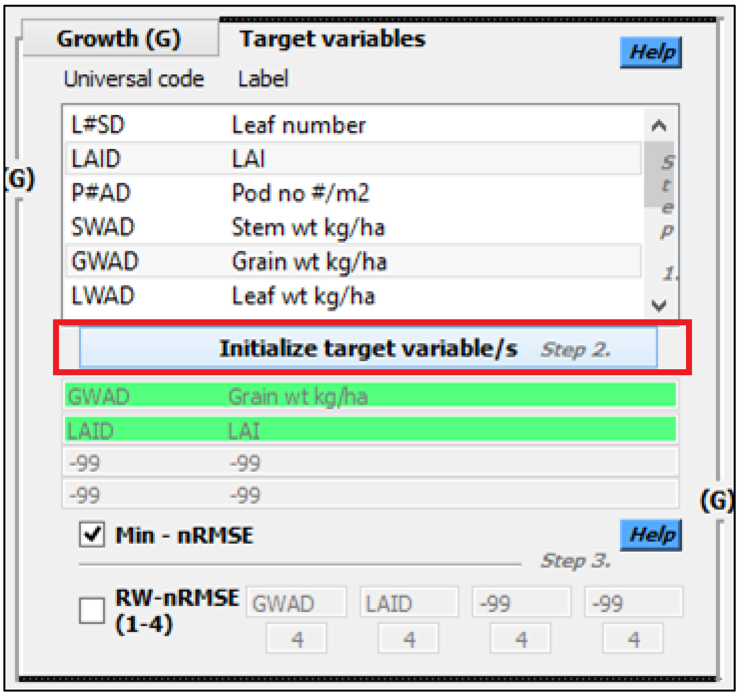


7.5 Interface – Help

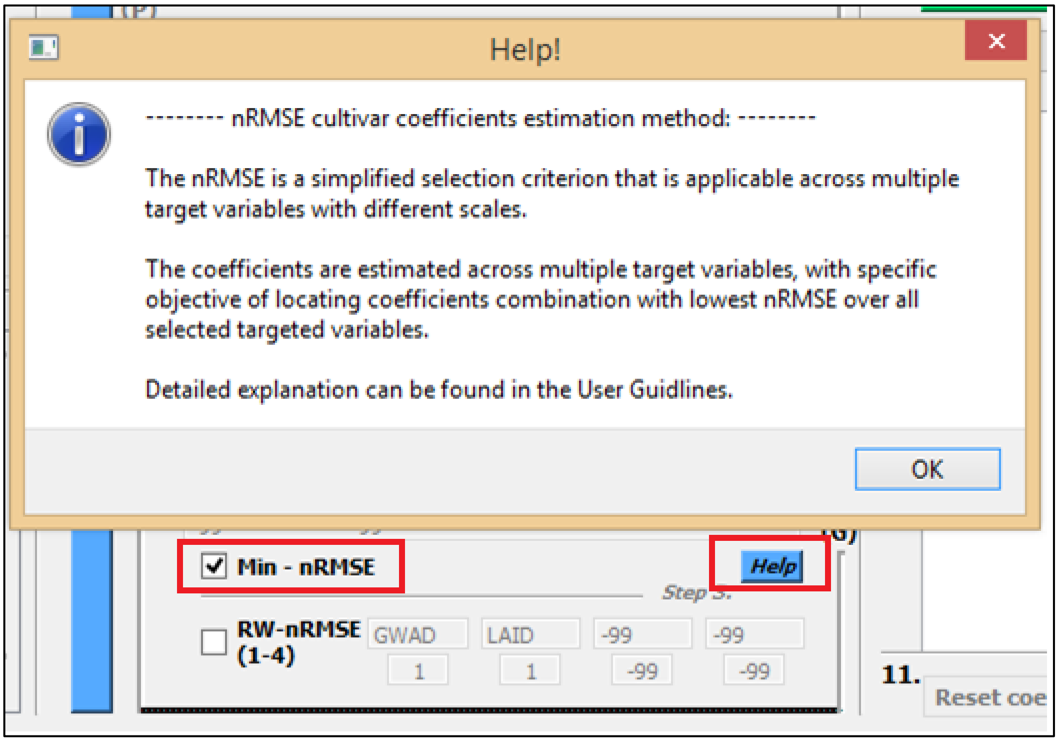
"Ctrl + mouse left click"- for selecting multiple target variables simultaneously!



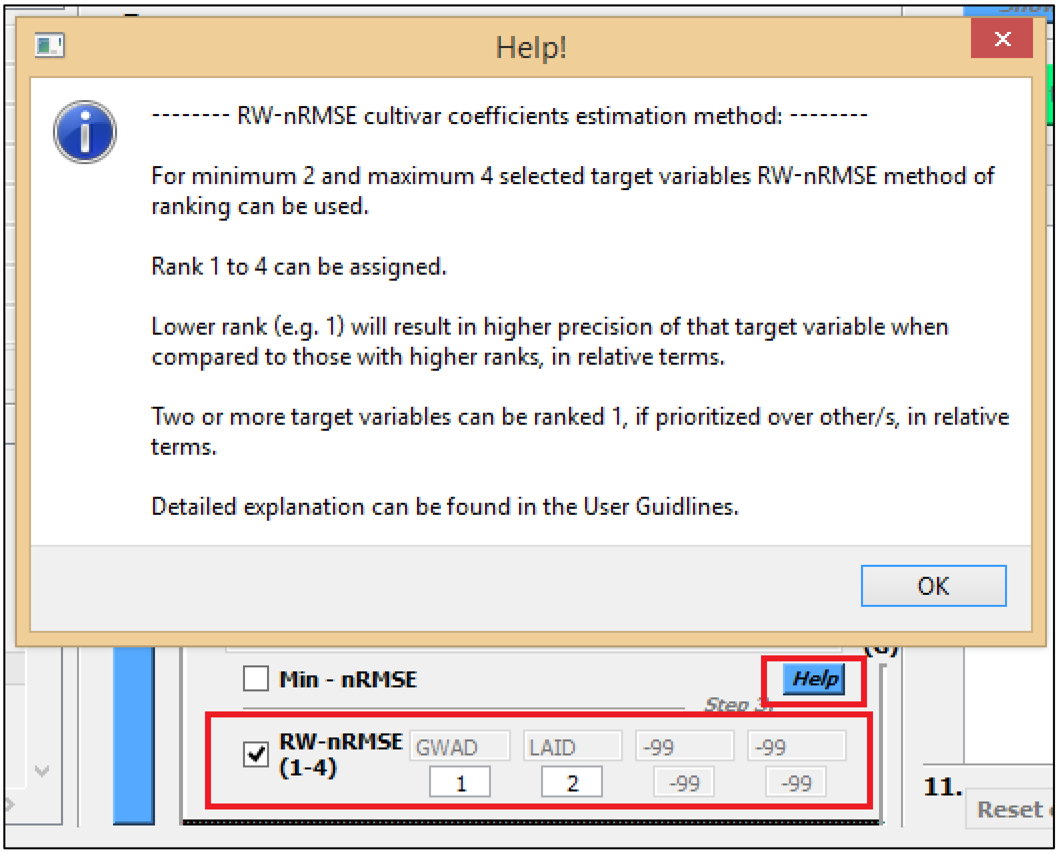
7.6 Interface



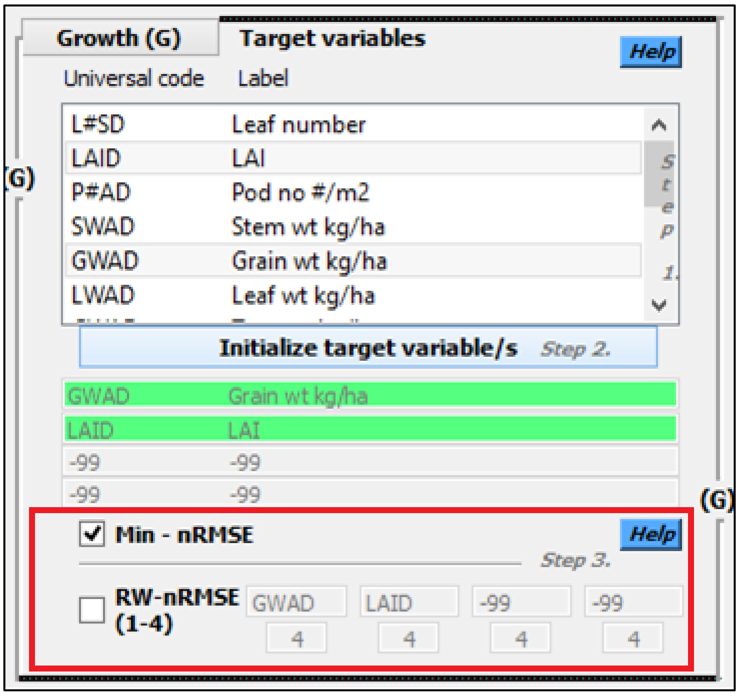
7.7 Interface



7.8 Interface – Help



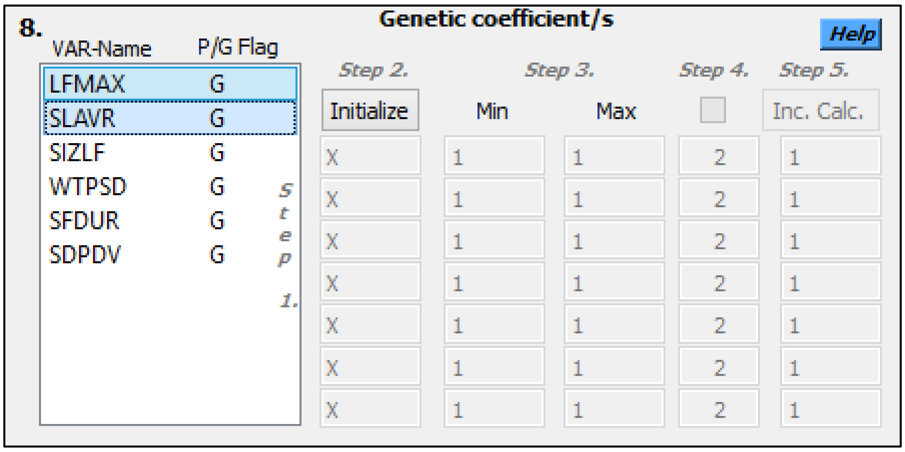
7.9 Interface – Help



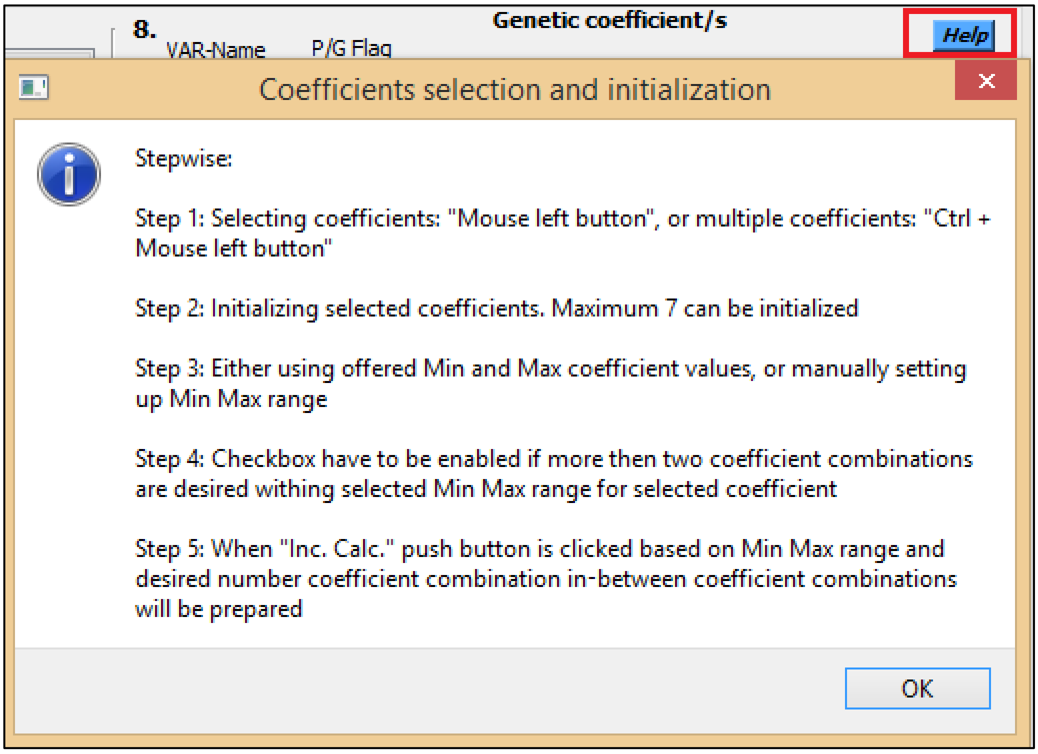
7.10 Interface

8. *Selecting desired coefficients and coefficient ranges and increment steps.*

First coefficient/s are selected from the list:

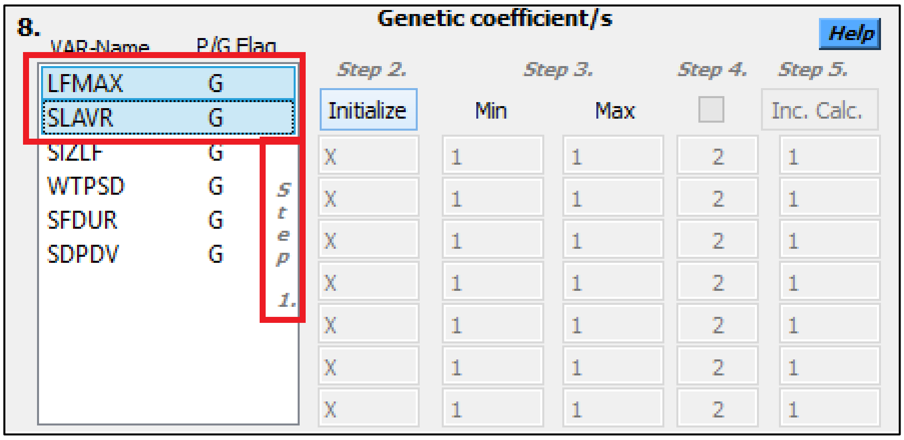


8.1 Interface

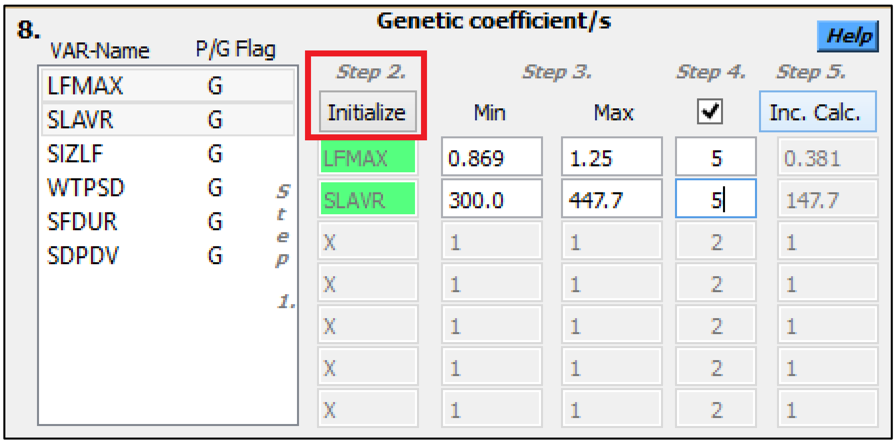


8.2 Interface - Help

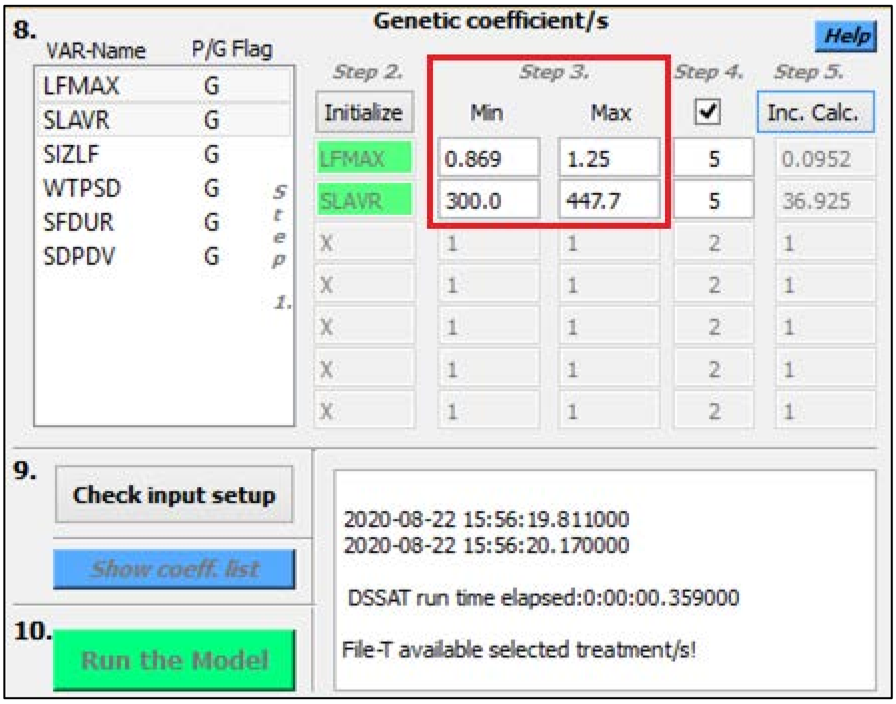
After selecting coefficient/s "Initialize" push button is clicked!



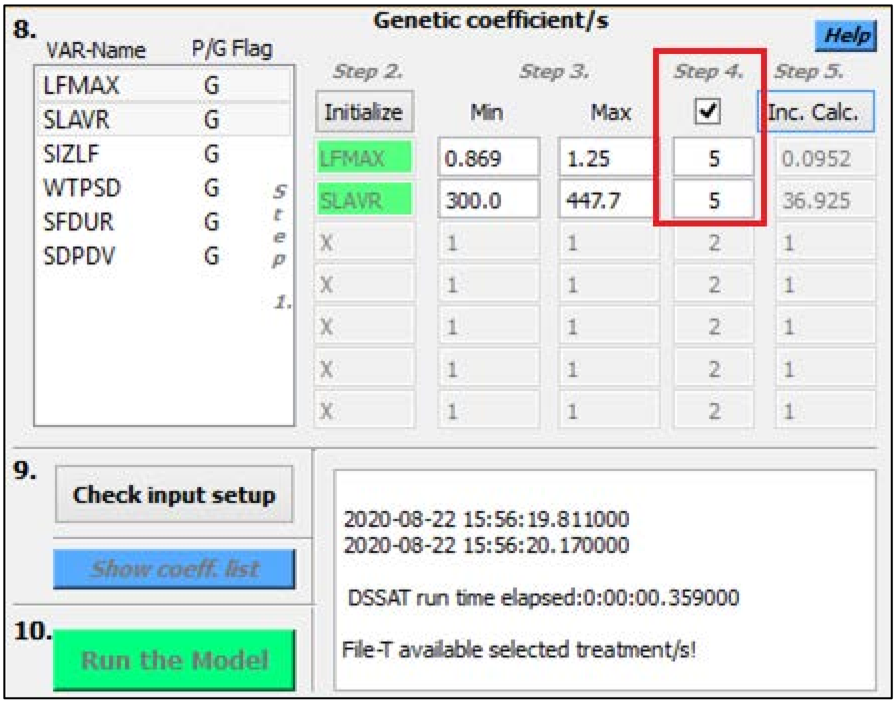
8.3 Interface



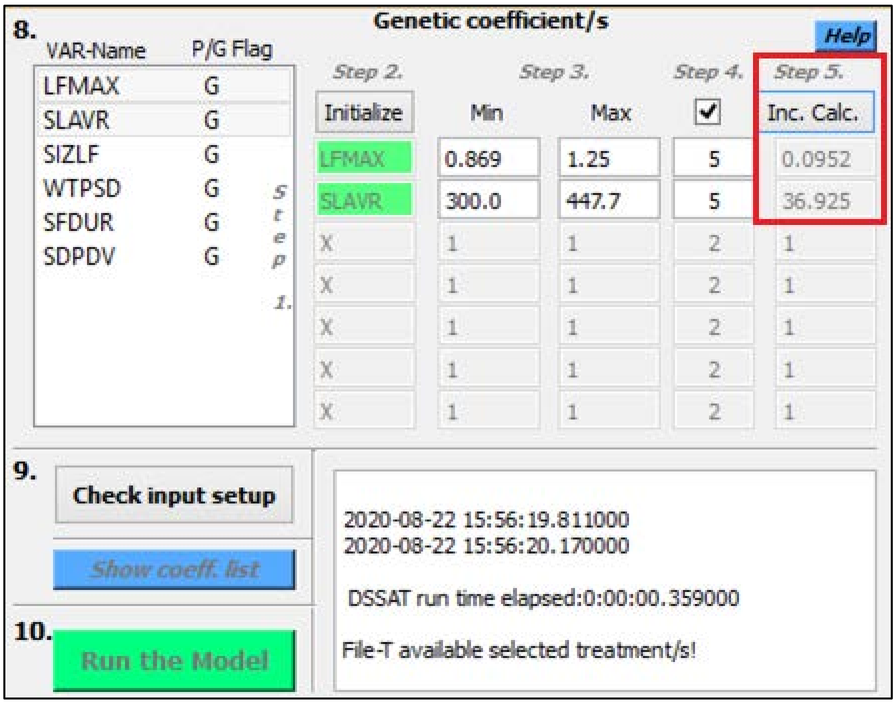
8.4 Interface



8.5 Interface



8.6 Interface

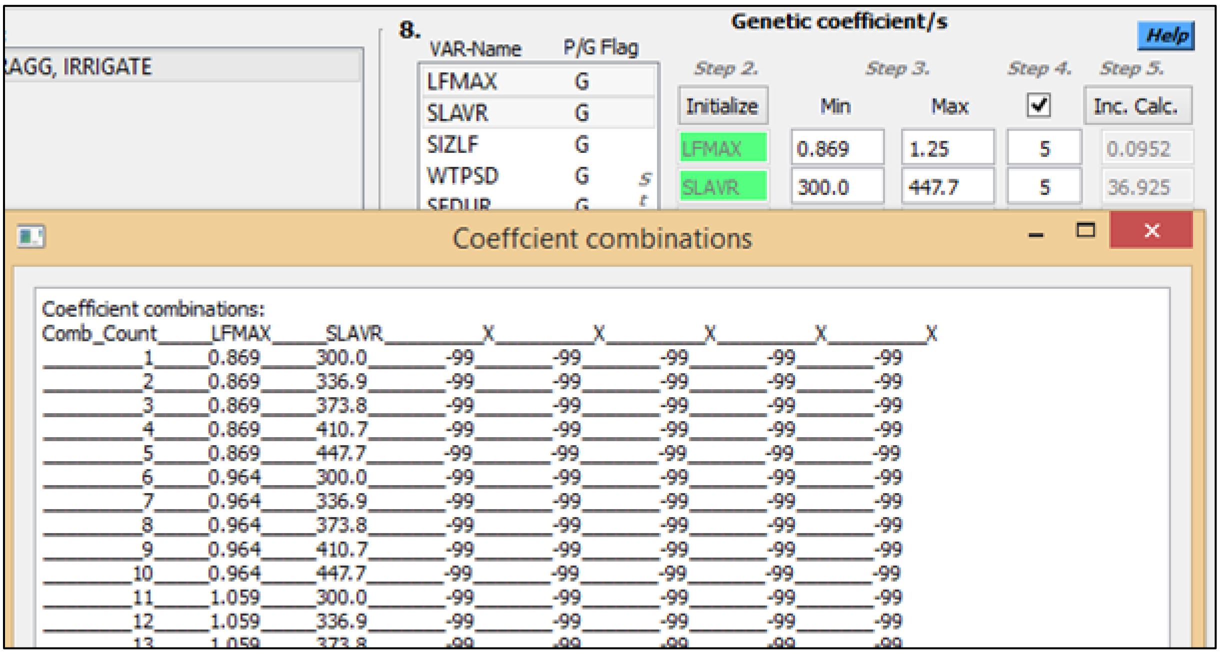


8.7 Interface

After selecting desired coefficients in section 8. from list widget window (step 1.) Initialize push button (step 2.) will initialize coefficient Labels in edit boxes and populate Min/Max coefficient ranges with min/max value available in corresponding cultivar file for selected cultivar coefficient (step 3.). Automatically this will create simple range of Min/Max value with two coefficient combinations that are going to be passed into the cultivar file, where after each model is executed. If the user wants more combinations in between given Min/Max range **Check Box** next to the **Inc. Calc**. has to be **Checked** (step 4.) and number of desired coefficient combinations between Min/Max can be given. After giving the desired number of combinations between Min/Max increment step is calculated by **Inc. Calc.** push button (step 5.) [(Max-

Min)/number of combinations]. If user would like to modify Min/Max values it can be done directly (manually) in edit boxes below Min/Max labels in step 2. If Min/Max are modified based on the number of combinations after Inc. Calc. push button is clicked new Min/Max and Inc. values are calculated and populated in corresponding edit boxes in section 8.

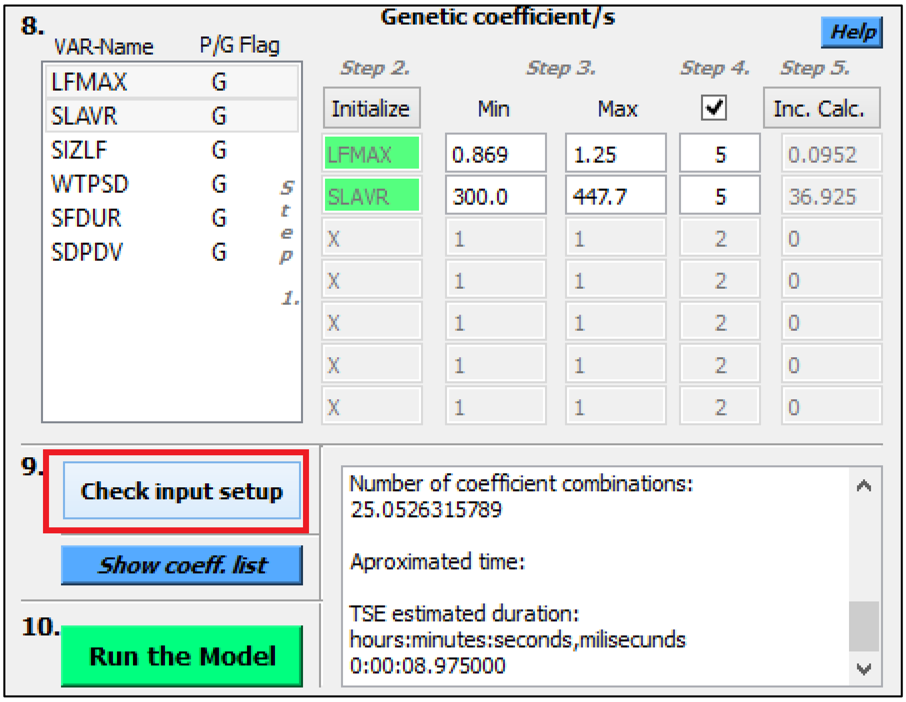
Example 1. With one coefficient (range and increment)



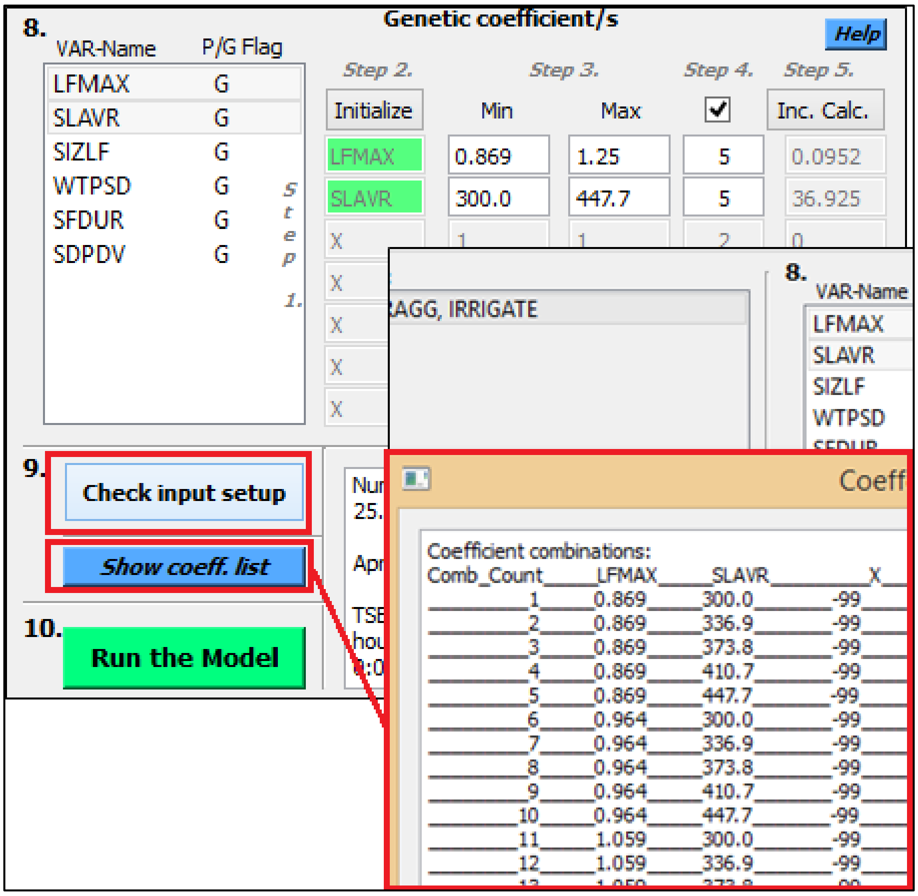
Example 2. With two coefficients (ranges and increments)

If the user wants to fix the value of some coefficient, in other words not to vary it during the program run, then they set Min and Max to equal value (<u>same value in the Min edit box as in the Max value edit box</u>).

9. Check if optimisation software setup is correct.

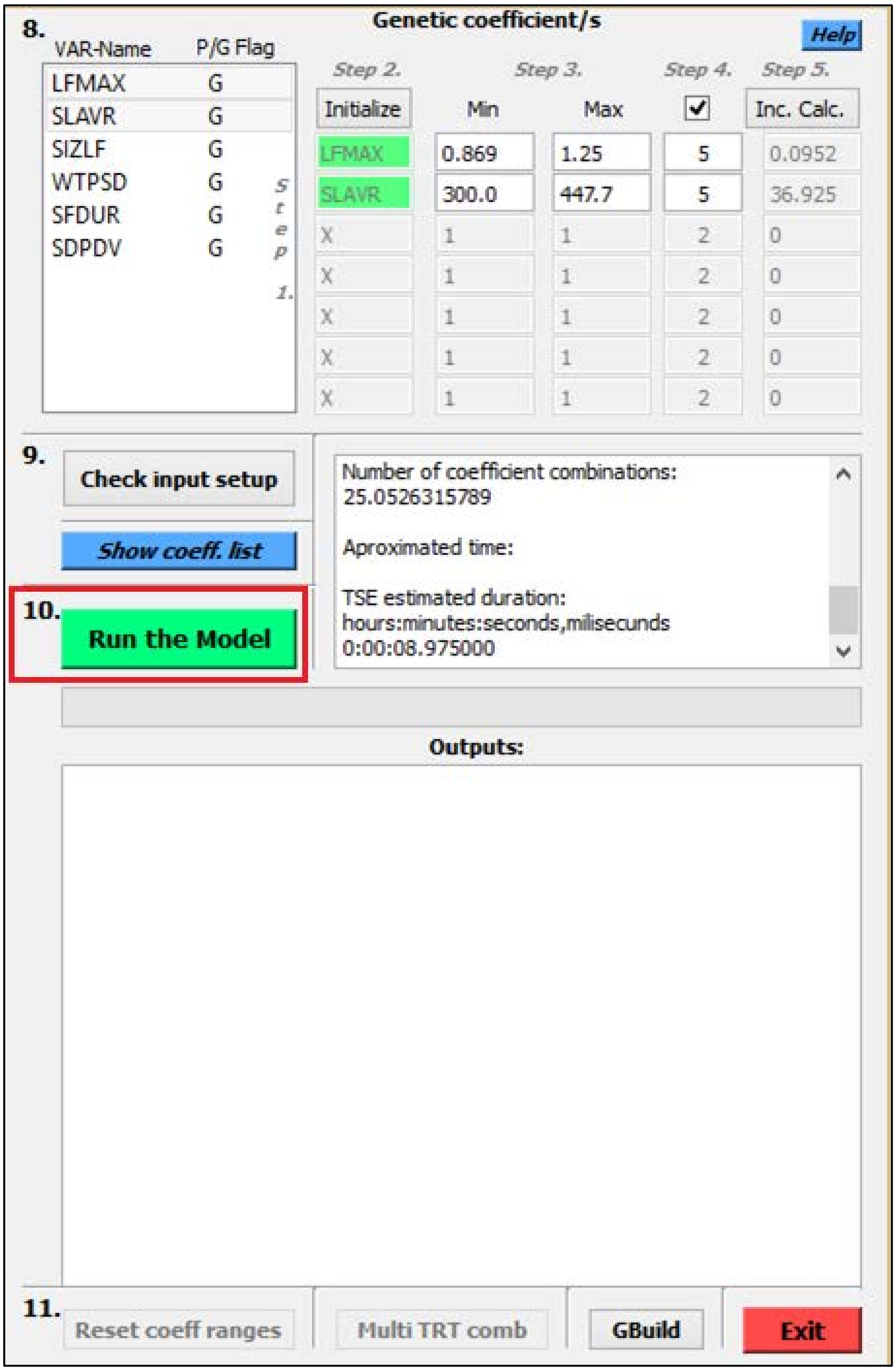


9.1 Interface

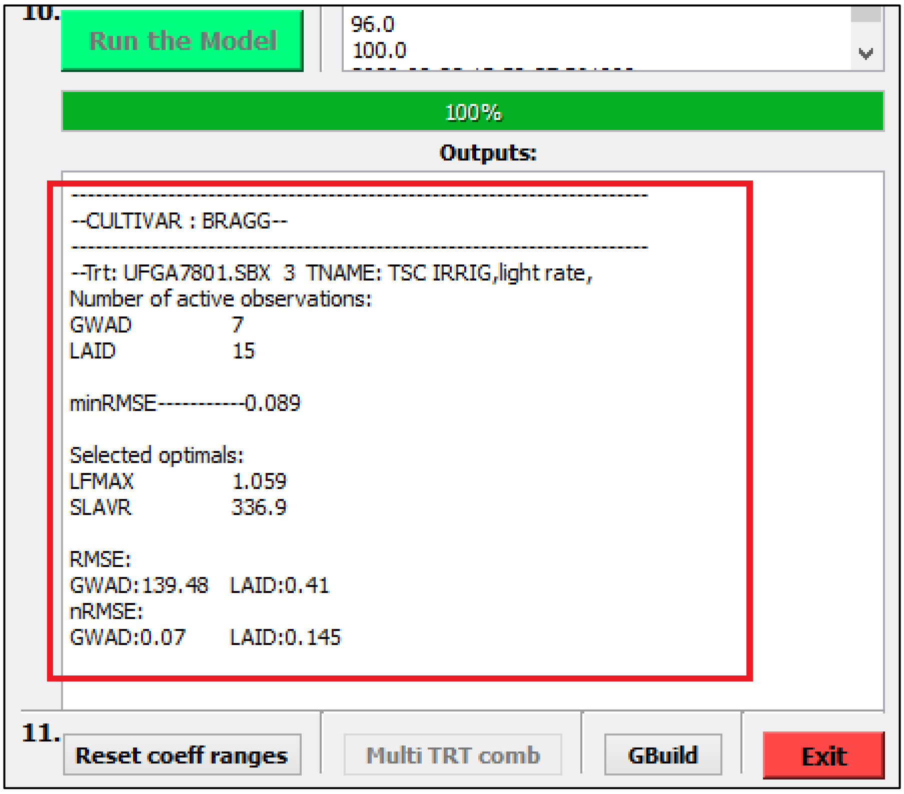


9.2 Interface

10. Run the model!



10.1 Interface

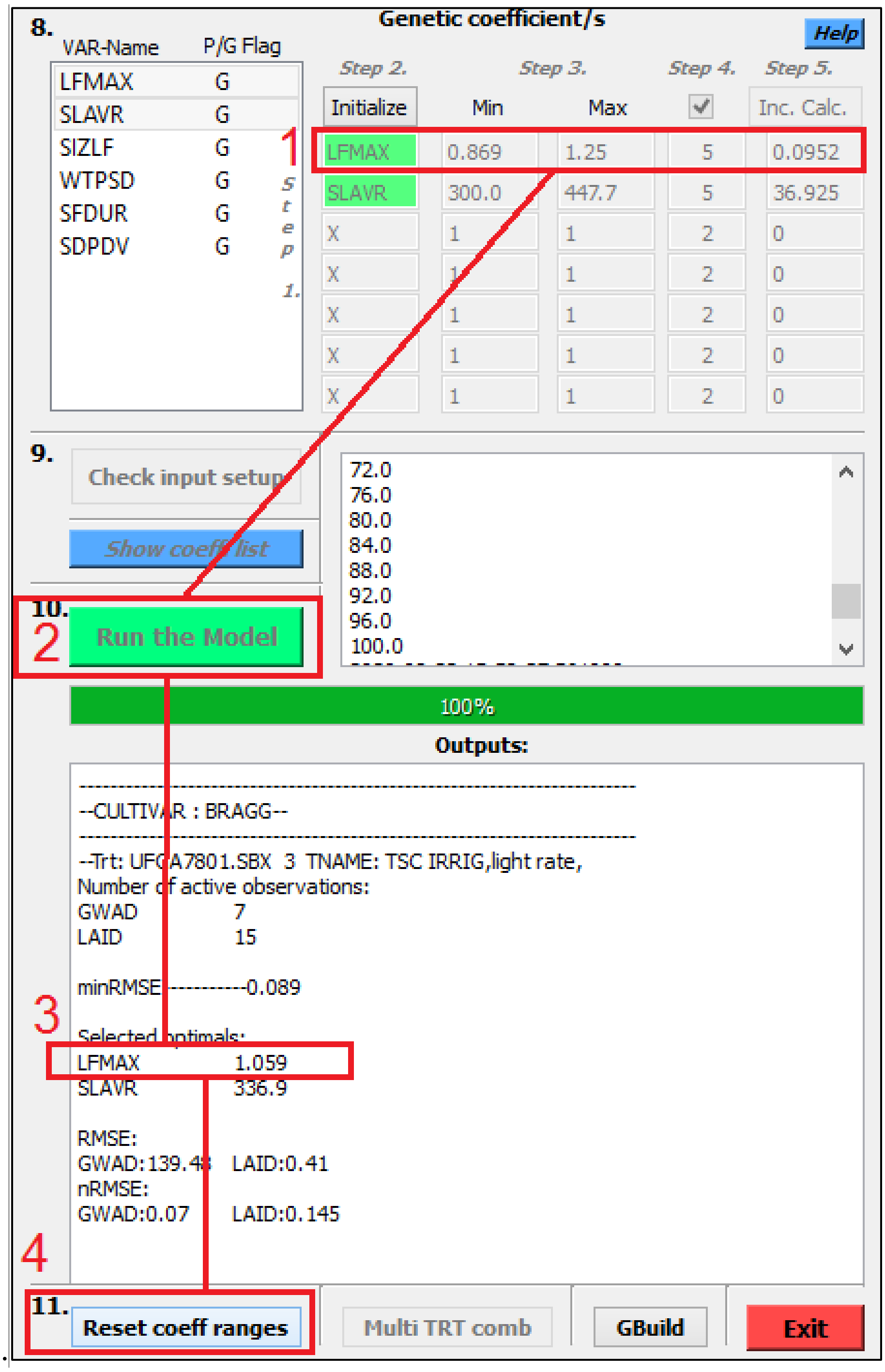


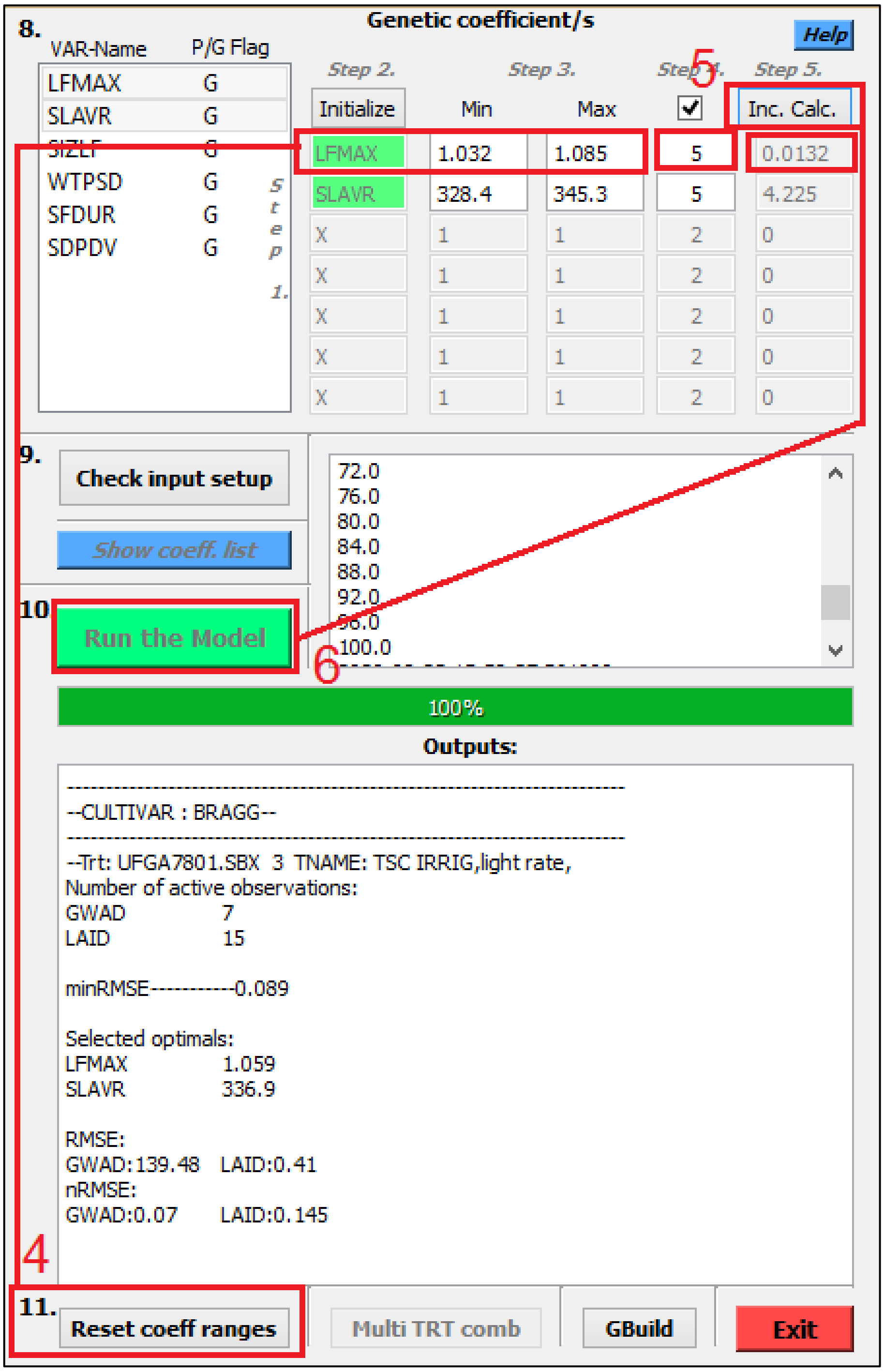
10.2 Interface

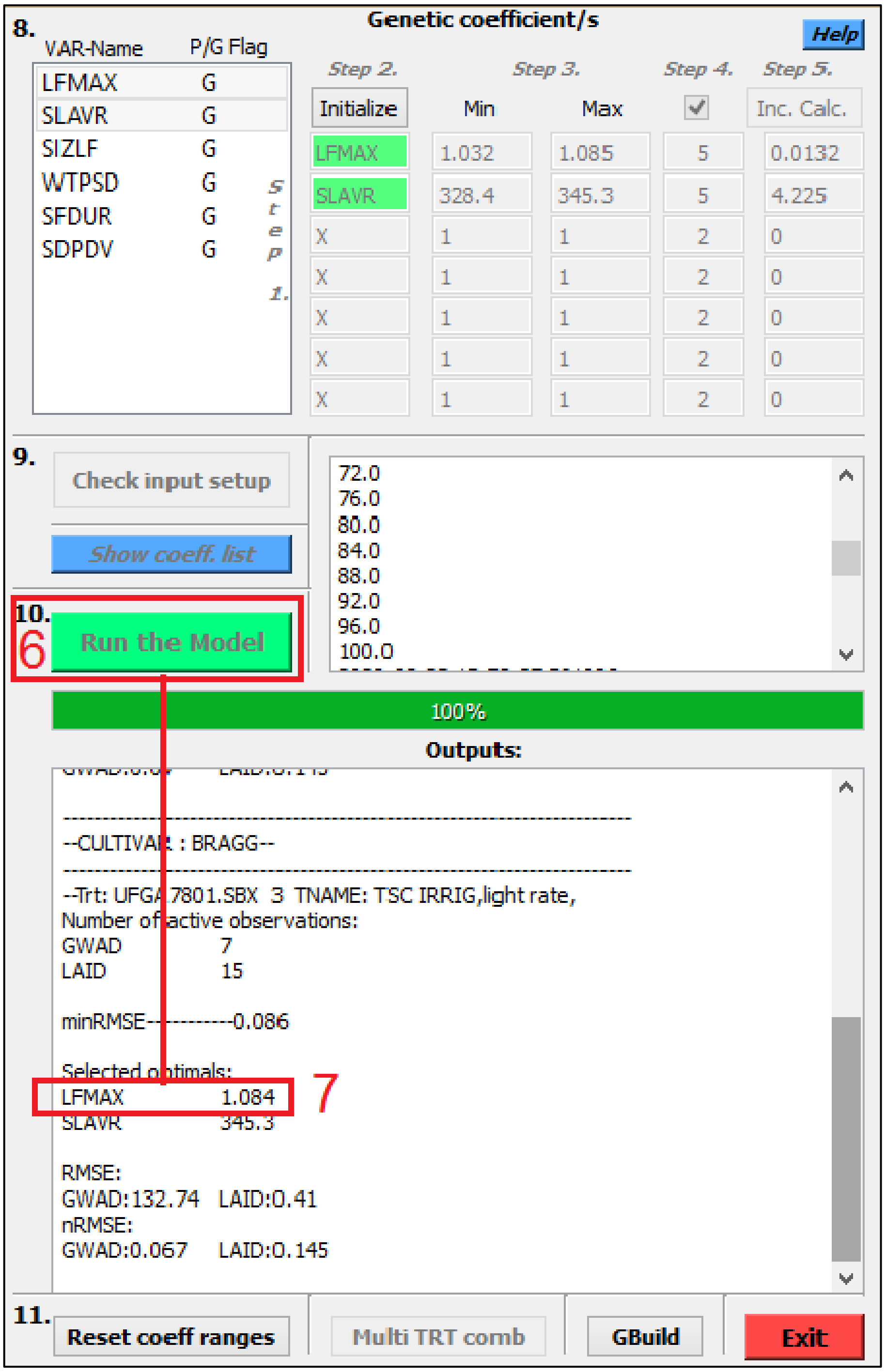
11. Additional options

I. [Reset coefficient ranges] or estimate

If “optimal” coefficient combination is found based on wide ranges and big increment steps this button will take those “optimums” and set new Min/Max range based on “optimum’s” values (“optimum” coefficient value * 0.25 and new Min=”optimum” – (“optimum”*0.25), Max=”optimum” + (“optimum”*0.25). Increment steps are calculated based on the desired number of combinations in between set in the section 8, step 4.



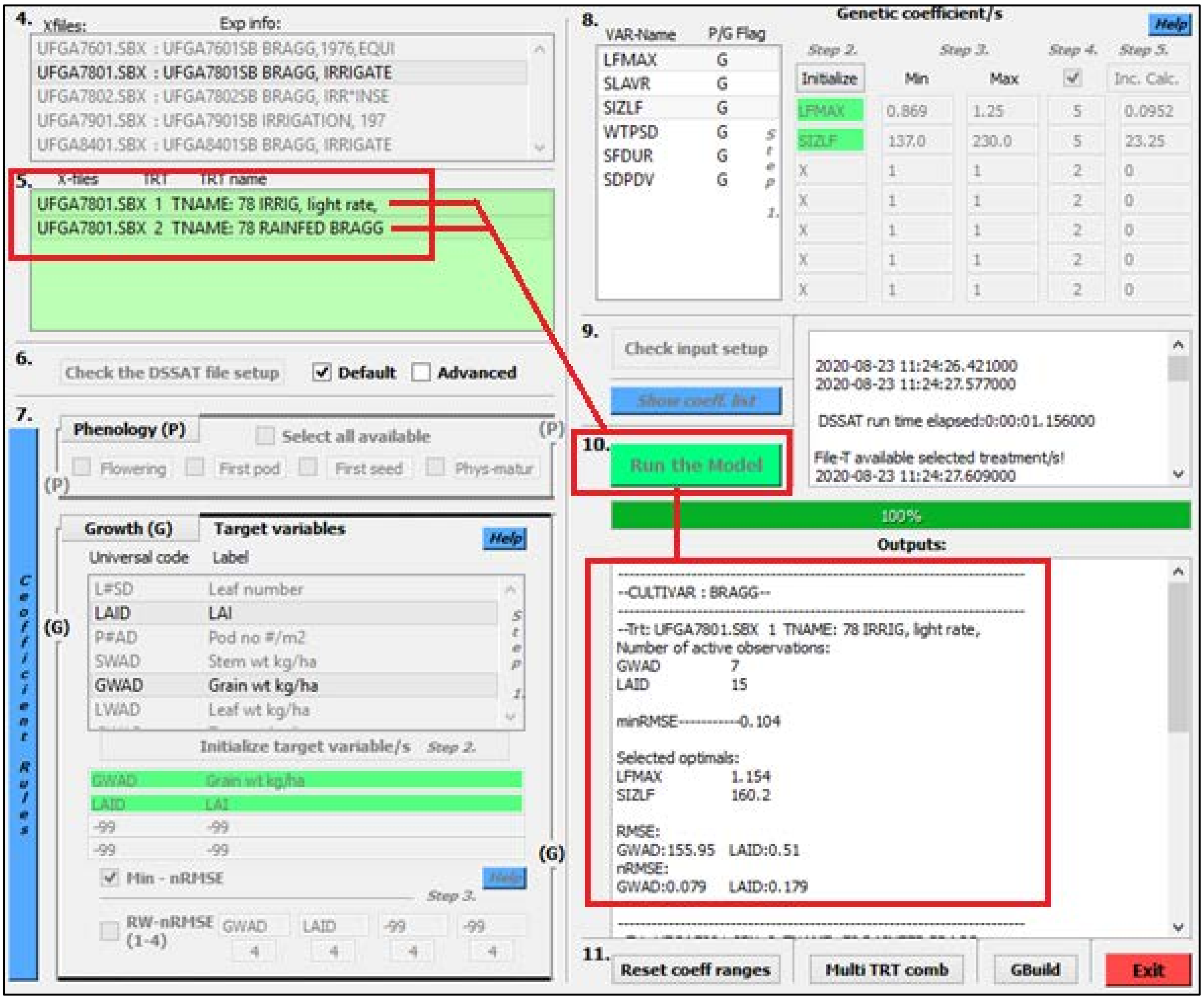

8.
Genetic coefficient/s
Help
VAR-Name
P/G Flag
LFMAX G
SLAVR G
SIZLF G
WTPSD G
SFDUR G
SDPDV G
Step 1.
Step 2.
Step 3.
Step 4.
Step 5.
Initialize
Min
Max
Inc. Calc.
LFMAX 1.032 1.085 5 0.0132
SLAVR 328.4 345.3 5 4.225
5
9.
Check input setup
Show coeff. list
10
Run the Model
6
72.0
76.0
80.0
84.0
88.0
92.0
100.0
100%
Outputs:
--CULTIVAR : BRAGG--
--Trt: UFGA7801.SBX 3 TNAME: TSC IRRIG,light rate,
Number of active observations:
GWAD 7
LAID 15
minRMSE-----------0.089
Selected optimals:
LFMAX 1.059
SLAVR 336.9
RMSE:
GWAD:139.48 LAID:0.41
nRMSE:
GWAD:0.07 LAID:0.145
4
11.
Reset coeff ranges
Multi TRT comb
GBuild
Exit

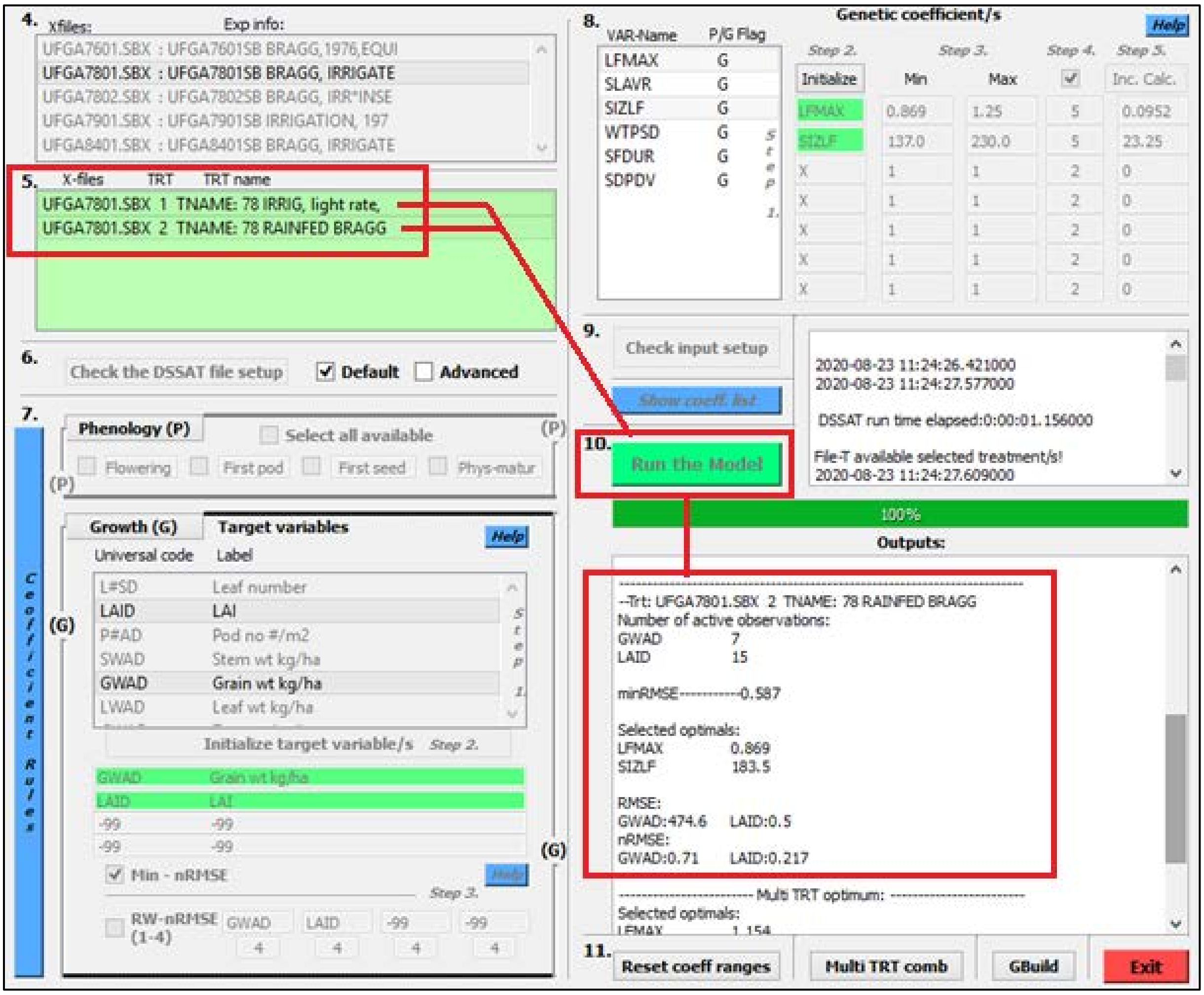


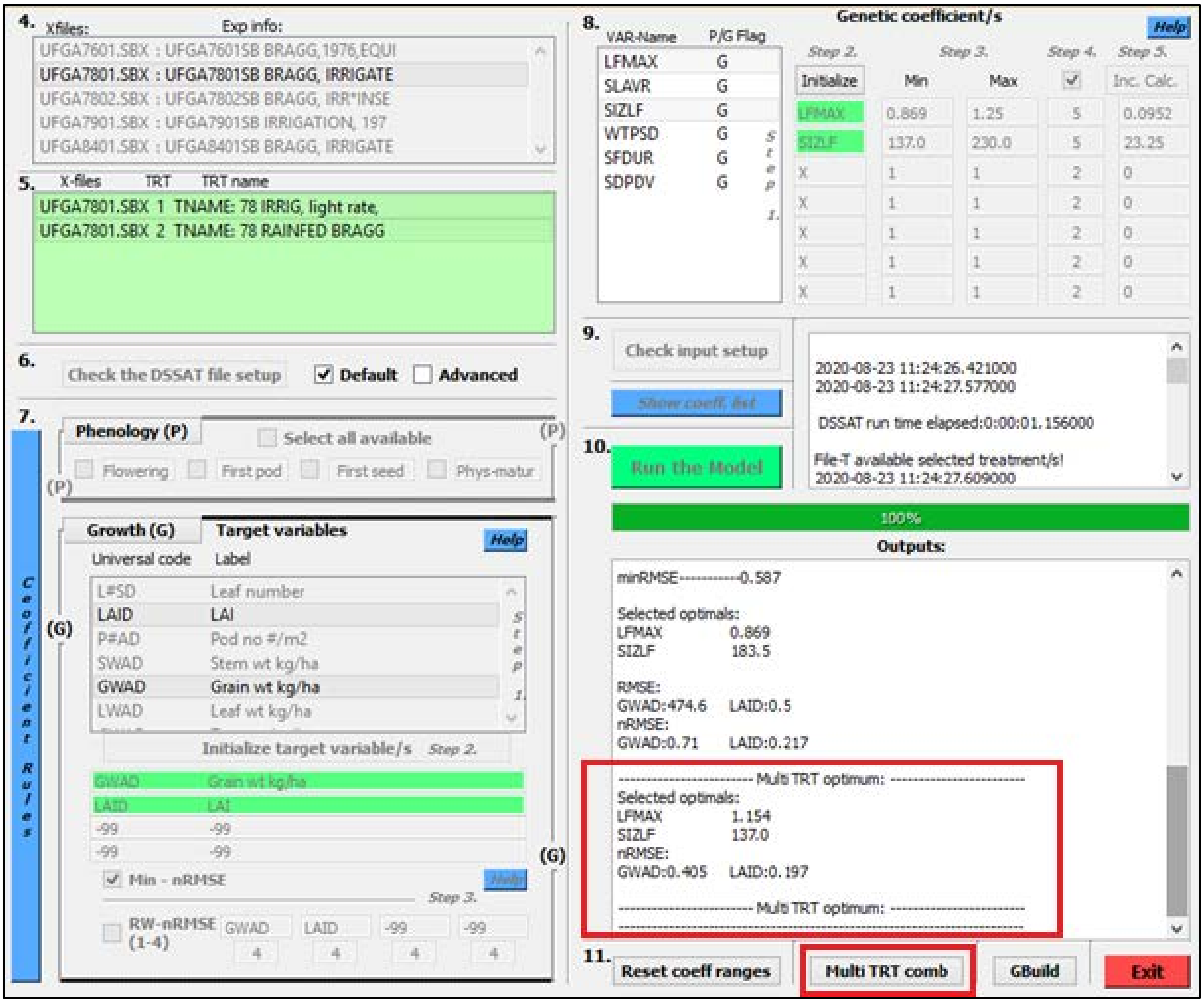
8.
Genetic coefficient/s
Help
VAR-Name
P/G Flag
LFMAX G
SLAVR G
SIZLF G
WTPSD G
SFDUR G
SDPDV G
Step 1.
Step 2.
Step 3.
Step 4.
Step 5.
Initialize
Min
Max
Inc. Calc.
LFMAX 1.032 1.085 5 0.0132
SLAVR 328.4 345.3 5 4.225
9.
Check input setup
Show coeff. list
72.0
76.0
80.0
84.0
88.0
92.0
96.0
100.0
10.
6
Run the Model
100%
Outputs:
--CULTIVAR : BRAGG--
--Trt: UFGA7801.SBX 3 TNAME: TSC IRRIG,light rate,
Number of active observations:
GWAD 7
LAID 15
minRMSE-----------0.086
Selected optimals:
LFMAX 1.084
SLAVR 345.3
7
RMSE:
GWAD:132.74 LAID:0.41
nRMSE:
GWAD:0.067 LAID:0.145
11.
Reset coeff ranges
Multi TRT comb
GBuild
Exit

II. [Multi TRT comb]

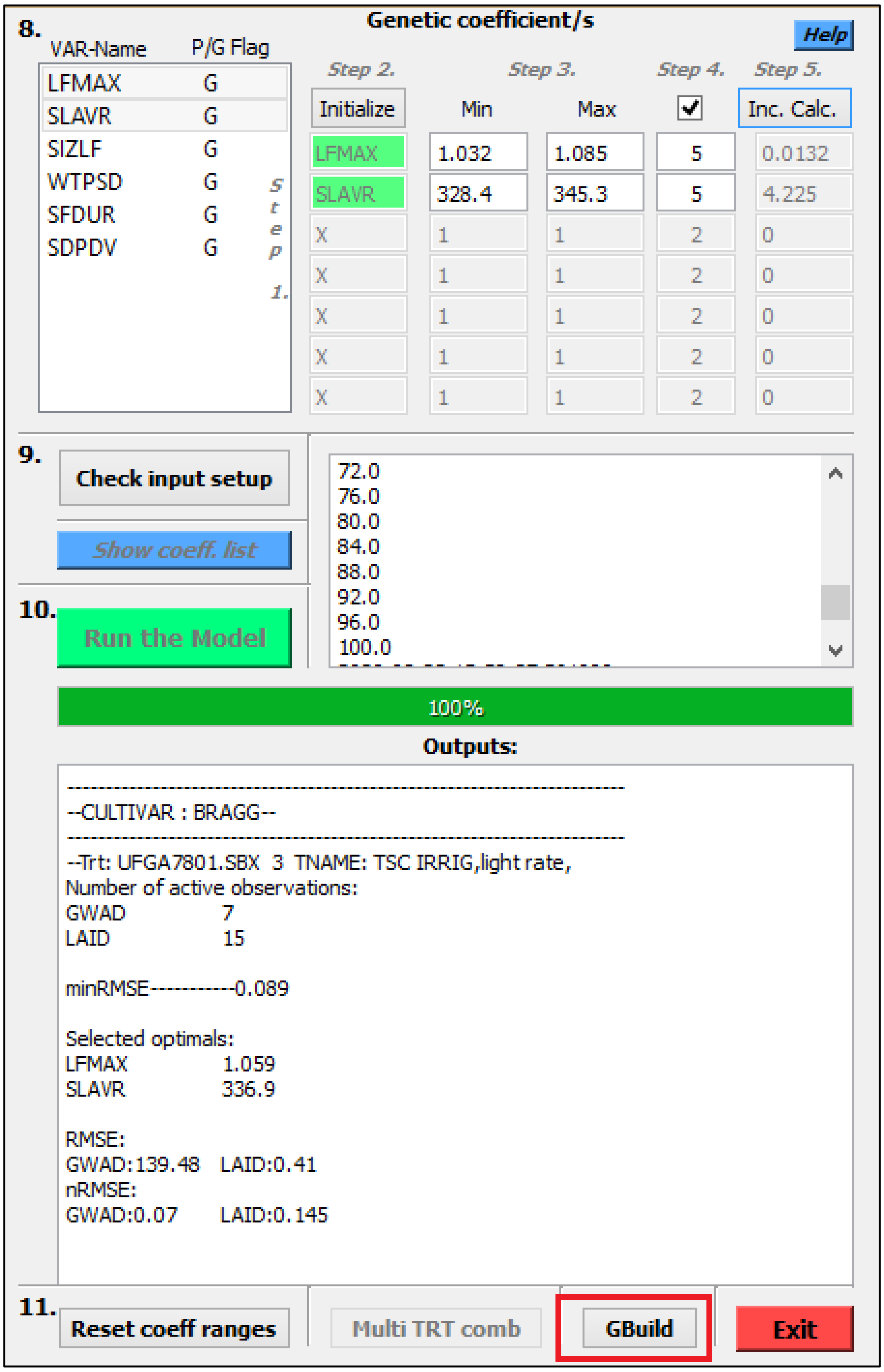
If this push button clicked multi treatment based cultivar coefficient combination will be estimated (as previously described).




4. Xfiles: Exp info:
UFGA7601.SBX : UFGA7601SB BRAGG,1976,EQUI
UFGA7801.SBX : UFGA7801SB BRAGG, IRRIGATE
UFGA7802.SBX : UFGA7802SB BRAGG, IRR*INSE
UFGA7901.SBX : UFGA7901SB IRRIGATION, 197
UFGA8401.SBX : UFGA8401SB BRAGG, IRRIGATE
5. X-files TRT TRT name
UFGA7801.SBX 1 TNAME: 78 IRRIG, light rate,
UFGA7801.SBX 2 TNAME: 78 RAINFED BRAGG
6. Check the DSSAT file setup
Default
Advanced
7. Phenology (P)
Select all available
Flowering
First pod
First seed
Phys-matur
Growth (G)
Target variables
Help
Universal code Label
L#SD Leaf number
LAID LAI
P#AD Pod no #/m2
SWAD Stem wt kg/ha
GWAD Grain wt kg/ha
LWAD Leaf wt kg/ha
Initialize target variable/s Step 2.
GWAD Grain wt kg/ha
LAID LAI
Min - nRMSE
Step 3.
RW-nRMSE (1-4)
Coefficient Rules
8. VAR-Name P/G Flag
LFMAX G
SLAVR G
SIZLF G
WTPSD G
SFDUR G
SDPDV G
Genetic coefficient/s
Step 2. Step 3. Step 4. Step 5.
Initialize Min Max Inc. Calc.
LFMAX 0.869 1.25 5 0.0952
SIZLF 137.0 230.0 5 23.25
9. Check input setup
Show coeff. list
2020-08-23 11:24:26.421000
2020-08-23 11:24:27.577000
DSSAT run time elapsed:0:00:01.156000
File-T available selected treatment/s!
2020-08-23 11:24:27.609000
10. Run the Model
100%
Outputs:
--Trt: UFGA7801.SBX 2 TNAME: 78 RAINFED BRAGG
Number of active observations:
GWAD 7
LAID 15
minRMSE-----------0.587
Selected optimals:
LFMAX 0.869
SIZLF 183.5
RMSE:
GWAD:474.6 LAID:0.5
nRMSE:
GWAD:0.71 LAID:0.217
Multi TRT optimum:
Selected optimals:
11. Reset coeff ranges
Multi TRT comb
GBuild
Exit

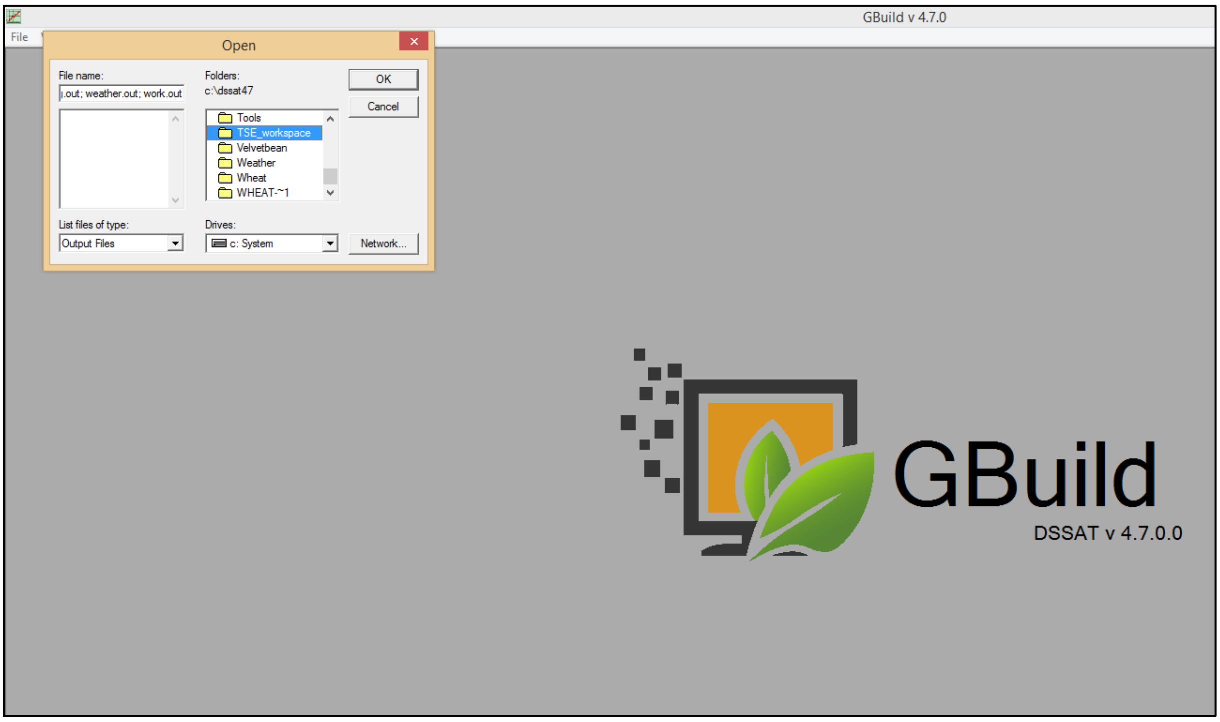

4. Xfiles: Exp info:
UFGA7601.SBX : UFGA7601SB BRAGG,1976,EQUI
UFGA7801.SBX : UFGA7801SB BRAGG, IRRIGATE
UFGA7802.SBX : UFGA7802SB BRAGG, IRR*INSE
UFGA7901.SBX : UFGA7901SB IRRIGATION, 197
UFGA8401.SBX : UFGA8401SB BRAGG, IRRIGATE
5. X-files TRT TRT name
UFGA7801.SBX 1 TNAME: 78 IRRIG, light rate,
UFGA7801.SBX 2 TNAME: 78 RAINFED BRAGG
6. Check the DSSAT file setup
Default
Advanced
7. Phenology (P)
Select all available
Flowering
First pod
First seed
Phys-matur
Growth (G)
Target variables
Help
Universal code Label
L#SD Leaf number
LAID LAI
P#AD Pod no #/m2
SWAD Stem wt kg/ha
GWAD Grain wt kg/ha
LWAD Leaf wt kg/ha
Initialize target variable/s Step 2.
GWAD Grain wt kg/ha
LAID LAI
Min - nRMSE
Step 3.
RW-nRMSE (1-4)
Coefficient Rules
8. VAR-Name P/G Flag
Genetic coefficient/s
LFMAX G
SLAVR G
SIZLF G
WTPSD G
SFDUR G
SDPDV G
Step 1.
Step 2. Step 3. Step 4. Step 5.
Initialize Min Max Inc. Calc.
LFMAX 0.869 1.25 5 0.0952
SIZLF 137.0 230.0 5 23.25
9. Check input setup
Show coeff. list
2020-08-23 11:24:26.421000
2020-08-23 11:24:27.577000
DSSAT run time elapsed:0:00:01.156000
File-T available selected treatment/s!
2020-08-23 11:24:27.609000
10. Run the Model
100%
Outputs:
minRMSE----------0.587
Selected optimals:
LFMAX 0.869
SIZLF 183.5
RMSE:
GWAD:474.6 LAID:0.5
nRMSE:
GWAD:0.71 LAID:0.217
Multi TRT optimum:
Selected optimals:
LFMAX 1.154
SIZLF 137.0
nRMSE:
GWAD:0.405 LAID:0.197
Multi TRT optimum:
11. Reset coeff ranges
Multi TRT comb
GBuild
Exit

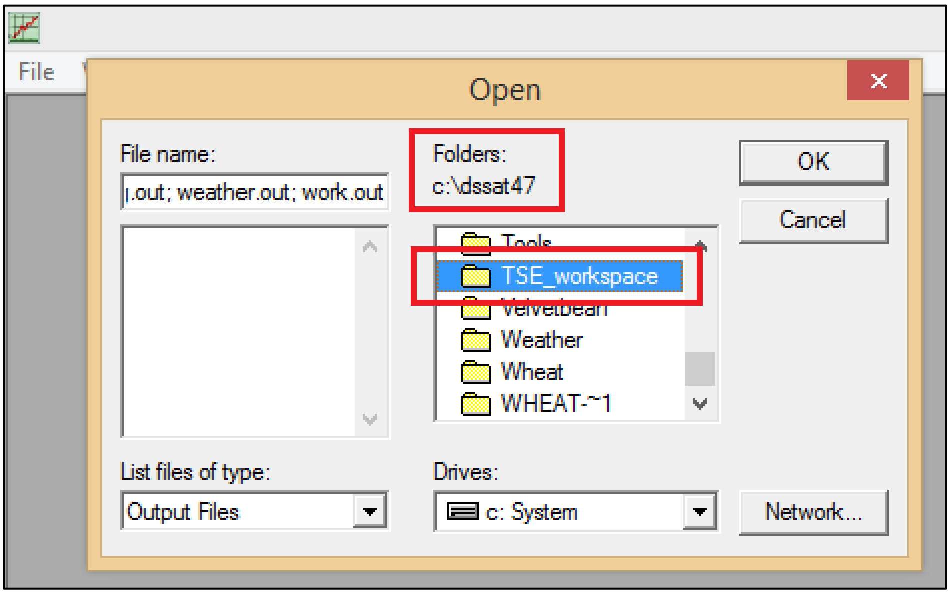
III. Execute [Gbuild]

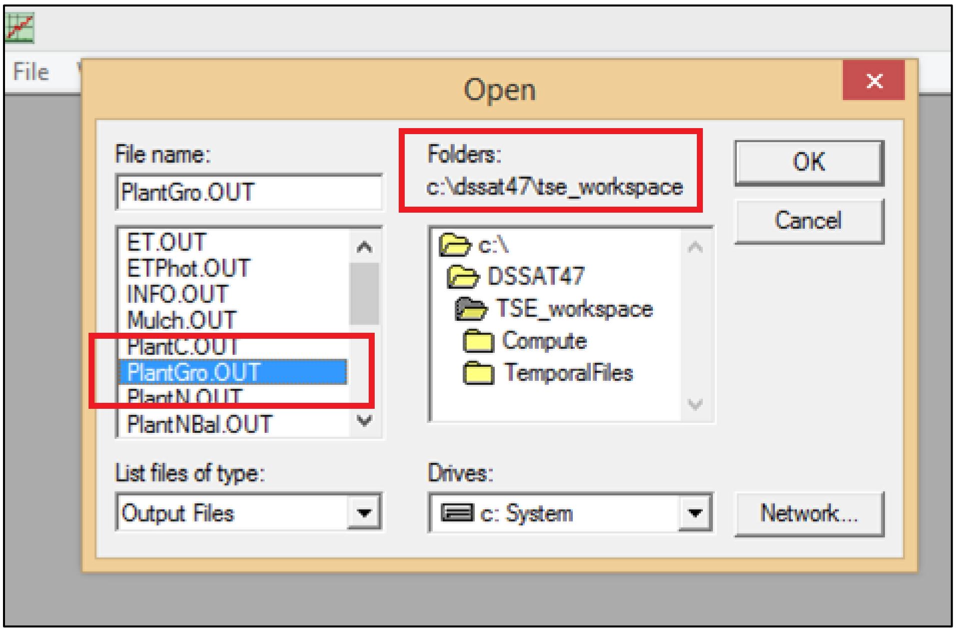


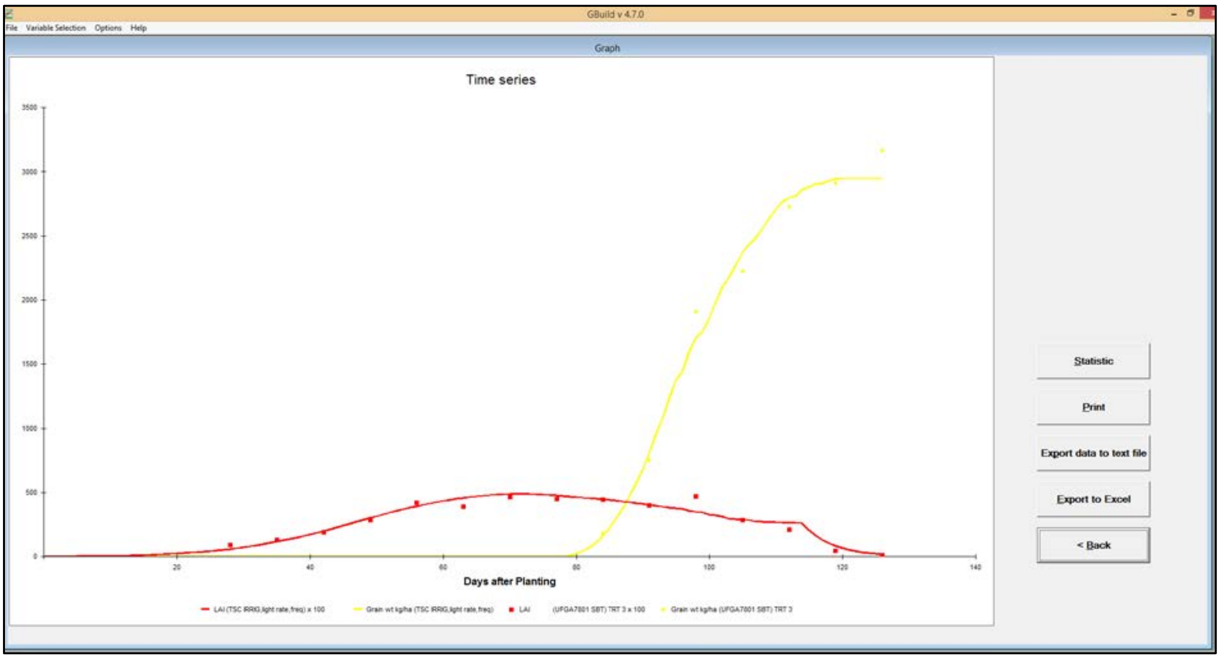
With **GBuild** PlantGro.Out can be opened **in TSE_workspace** directory to see how good new combination is. If more than one experiments were used (if Multi-TRT) combination was found, all used treatments can be seen in Gbuid simultaneously…

Again if the PlantGro.Out in TSE_workspace is open with GBuild visual and statistical fit across all used experiments of coefficient optimisation can be seen in GBuild.


GBuild v 4.7.0
File
Open
File name:
|.out; weather.out; work.out
Folders:
c:\dssat47
OK
Cancel
Tools
TSE_workspace
Velvetbean
Weather
Wheat
WHEAT-~1
List files of type:
Output Files
Drives:
c: System
Network...
GBuild
DSSAT v 4.7.0.0


File
Open
File name:
|.out; weather.out; work.out
Folders:
c:\dssat47
OK
Cancel
Tools
TSE_workspace
Velvetbean
Weather
Wheat
WHEAT-~1
List files of type:
Output Files
Drives:
c: System
Network...





IV. **[Exit] the program and all running threads**

APPENDIX

1 The nRMSE error minimisation method

The *nRMSE* is a simplified selection criterion that is applicable across multiple target variables with different scales. The coefficients are estimated across multiple target variables, with specific objective of locating coefficients combination with lowest *nRMSE* over all targeted variables (Eq. 1).

$$AVG_{nRMSE(n)} = (GWAD_{nRMSE(n)} + LAID_{nRMSE(n)} + CWAD_{nRMSE(n)} + LWAD_{nRMSE(n)})/4 \quad (1)$$

The selection of the coefficient combination with the lowest nRMSEs average across all target variables proves to be a good solution (Table 4, AVG-nRMSE=0.12).

Table *4* The *nRMSE* - simplified example of varying one of the cultivar parameters affecting growth (G) related target variables (GWAD, LAID, CWAD, and LWAD) of Gainesville 1978 treatment with calculated nRMSEs for multiple target variables (each) and optimum selection based on the average nRMSE (AVG-nRMSE) over multiple target variables (Memic et al.2020).

LFMAX	**nRMSE**				**AVG**	
(G -Growth**)**	**GWAD**	**LAID**	**CWAD**	**LWAD**	**nRMSE**	
0.8	0.208	0.22	0.185	0.203	0.204	
0.912	0.131	0.153	0.119	0.146	0.137	
1.024	**0.082**	0.145	0.109	0.144	**0.12**	**0.12**
1.136	0.078	0.171	0.137	0.173	0.14	
1.248	0.109	0.209	0.178	0.211	0.177	

LFMAX – CROPGRO coeff., *GWAD* - grain weight, *LAID* - leaf area index, *CWAD* - tops weight, *LWAD* - leaf weight

2 The nRMSE multiple treatment based goodness of fit criteria

Cultivar coefficients can be optimised based on the in-season observation of one experiment (single treatment/experiment) or multiple experiment data sets (multiple season and/or locations). Using only one experimental data set for optimisation will lead to better statistical agreement between simulated and observed, but potentially result in over-fitting of simulation outputs with observations. Cultivar coefficients should perform well across multiple seasons and locations and as such should be derived based on multiple season and locations. In Table 5 are shown single treatment "optimums" and multiple treatment "optimum".

For demonstrating single treatment and multiple treatment based cultivar coefficient optimisation one cultivar coefficient (LFMAX) and three treatments have been selected (Gainesville 1978, Gainesville 1979 and Gainesville 1984). Cultivar coefficient value minimum (0.85), maximum (1.25) and increment step (0.1) were written in the corresponding cultivar file and crop model was executed. For each cultivar coefficient simulated and observed GWAD, LAID, CWAD and LWAD average nRMSE was calculated for localising "optimums". First single treatment "optimums" are localised based on the lowest average nRMSE (AVG-nRMSE, Table 5, section a, grey fields). In second step multiple treatment cultivar coefficient "optimum" is localised based on the single treatment based AVG-nRMSEs with formula:

$$([TRT_1+TRT_2+TRT_3]/3)$$

Based on the averaging of the single treatment based AVG-nRMSEs multiple treatment based optimum is selected (Table 5, section b, grey fields) with lowest treatment based average nRMSE (Table 5, section b, blue field).

Table 5 'Bragg' cultivar (*DSSAT Default*) is shown for soybean experiments conducted in Gainesville 1978, Gainesville 1979 and Gainesville 1984, with LFMAX cultivar coefficient variations from 0.85 to 1.25 with increment step 0.1 with 5 coefficient combinations for each treatment.

a) Single treatment "optimums"				b) **Multi treatment based "optimum"**					
Year	TRT	LFMAX	AVG nRMSE	Year	TRT	LFMAX	AVG nRMSE	Multiple treatment average	Optimum
1978	1	0.85	0.17	1978	1	0.85	0.17		
1978	1	0.95	0.128	1979	2	0.85	0.119		
1978	1	1.05	0.122	1984	3	0.85	0.408	0.232	
1978	1	1.15	0.142	1978	1	**0.95**	0.128		
1978	1	1.25	0.175	1979	2	**0.95**	0.14		
1979	2	0.85	0.119	1984	3	**0.95**	0.374	**0.214**	**0.214**
1979	2	0.95	0.14	1978	1	1.05	0.122		
1979	2	1.05	0.177	1979	2	1.05	0.177		
1979	2	1.15	0.21	1984	3	1.05	0.355	0.218	
1979	2	1.25	0.239	1978	1	1.15	0.142		
1984	3	0.85	0.408	1979	2	1.15	0.21		
1984	3	0.95	0.374	1984	3	1.15	0.346	0.233	
1984	3	1.05	0.355	1978	1	1.25	0.175		
1984	3	1.15	0.346	1979	2	1.25	0.239		
1984	3	1.25	0.345	1984	3	1.25	0.345	0.253	

- *LFMAX*-defined in Table 3,
- *TRT*-treatment,
- *AVG-nRMSE*-average of normalised RMSE over four target variables (grain weight, leaf area index, Tops weight, Leaf weight)

3 The RW-nRMSE

A *Relative Weight -nRMSE* (*RW-nRMSE*) complementary method was developed with priority ranks (1 to 4). A rank can be assigned to each target variable separately, where 1 is the most accurate and 4 the least accurate, in relative terms. Each accuracy rank has weight coefficient assigned to it in the TSE code (Table 6), rank 1 with weight coefficient of 0.01 and rank 4 with values 0.04 used in selection criteria. With ranks each target variable's accuracy can be selected, giving users more flexibility in terms of target variable prioritisation in the cultivar coefficients estimation process (Table 6), as shown in the following example with gwad having rank 1 and the other target variables rank 4.

Table 6 RW-nRMSE method target variable ranking with weight coefficients

RW-nRMSE		
	Rank	**Coefficient weight**
LAID	4	0.04
CWAD	4	0.04
LWAD	4	0.04
GWAD	**1**	**0.01**

Initially the weight coefficients are defined (by ranking) for creating a mathematical threshold used in the program (Figure 2a). The initial selection criteria was set to 0.0 (Figure 2a). With the initial selection criteria the *while* loop (flow controlling statement executed repeatedly until specified conditions are meet in the programming language) is initialised (Figure 2b). Within the *while* loop, the *if* condition is set for finding an overall statistically acceptable solution (Figure 2c). The *while* loop will be stopped only if the calculated nRMSEs for multiple target variables satisfies the RW-nRMSE selection criteria.

If accurate grain yield is the goal (other target variables less accurate) of the optimisation then the relative weight coefficients based on RW-nRMSE criteria (Table 6) gwad rank 1 and the other three variables rank 4 offers more accurate results of gwad target variable. The rank 1 for gwad and rank 4 for laid, cwad and lwad initializes RW-nRMSE mathematical threshold in to code for while loop first run of 0.01 for nRMSE-gwad and 0.04 for nRMSE-laid-cwad-lwad. In practice this means when the first loop run is searching through calculated nRMSEs for all target variables only coefficient combinations that have calculated nRMSE-gwad lower than 0.01 and nRMSE-laid-cwad-lwad lower than 0.04 is selected as optimum. If in first while loop run none of the coefficient combinations have calculated nRMSE as defined with relative weights the while loop is restarted for the second time with new mathematical threshold for nRMSE-gwad lower than 0.02 and nRMSE-laid-cwad-lwad lower than 0.8. If the newly formed mathematical threshold doesn't provide coefficient combination with calculated nRMSEs the while loop will be started again with nRMSE-gwad mathematical threshold of 0.03 etc. until coefficient combination providing calculated nRMSEs for target variables is located.

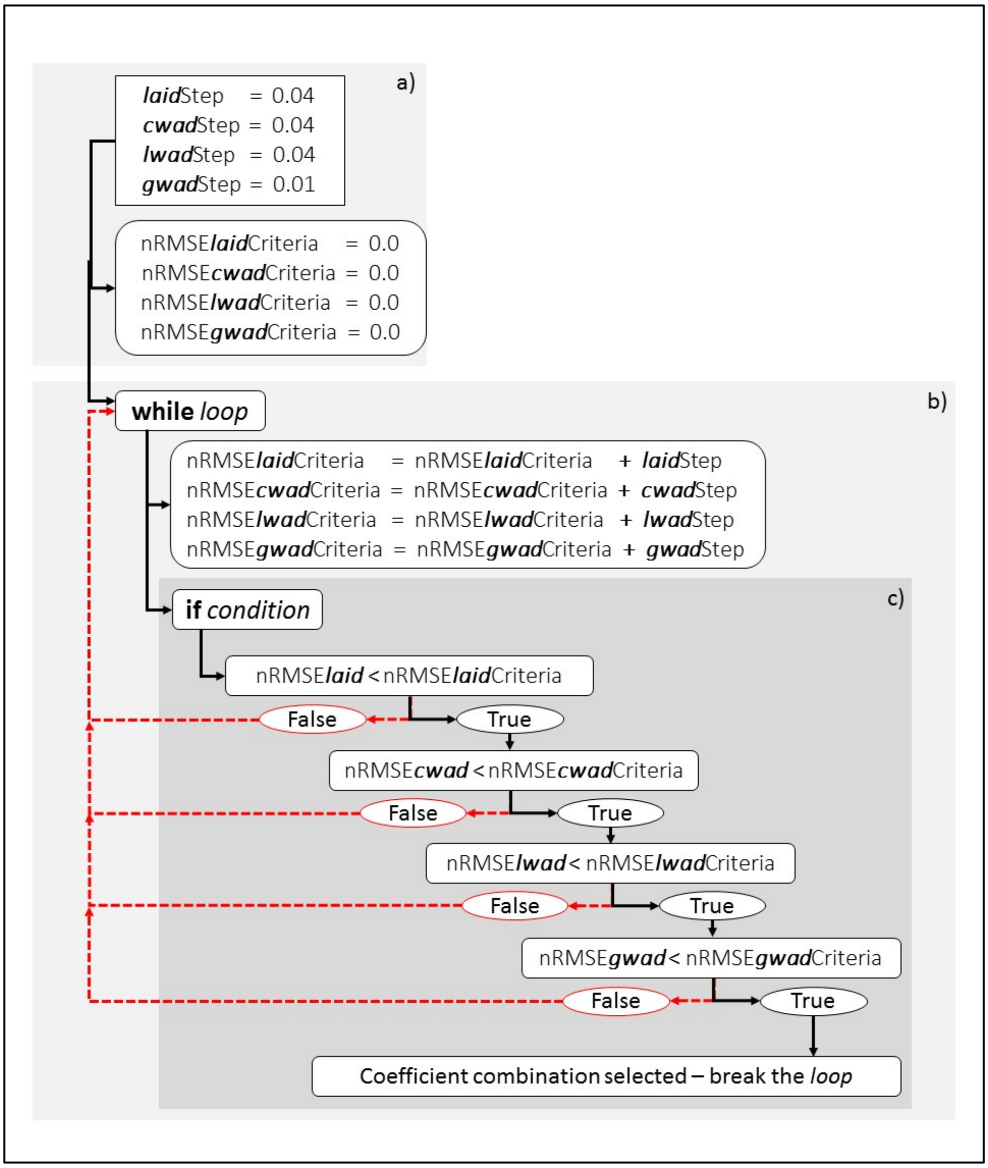


Figure 2 The RW-nRMSE cultivar coefficient combination selection method

With RW-nRMSE-gwad rank 1 and RW-nRMSE-laid-cwad-lwad rank 4 while loop with if conditions is repeated 8 times (Table 7) with desired LFMAX coefficient of 1.136 providing the most accurate simulations for grain weight (gwad). This coefficient values is different from the one selected with simple nRMSE criteria results shown in Table 5.

Table *7* The *RW-nRMSE* - simplified example of varying one of the cultivar parameters with calculated nRMSEs for multiple target variables demonstrating selected combinations with target variable with relative weights ranking in TSE.

	nRMSE				
LFMAX (G)	**gwad**	**laid**	**cwad**	**lwad**	**AVG-nRMSE**
0.8	0.208	0.22	0.185	0.203	0.204
0.912	0.131	0.153	0.119	0.146	0.137
1.024	0.082	0.145	0.109	0.144	**0.12**
1.136	**0.078**	0.171	0.137	0.173	0.14
1.248	0.109	0.209	0.178	0.211	0.177

G – *growth*, gwad - *grain weight*, laid - *leaf area index*, cwad - *tops weight*, lwad - *leaf weight*

4 Range reduction – generating coefficient combinations *(direct take out from paper: Memic at al. 2020 – in submission process)*

The cultivar coefficient estimation process consists of varying the values for each cultivar coefficient and comparing a statistical fit of simulated outputs with field observations in order to determine the coefficient combination providing the best agreement between simulated and observed values. Various cultivar coefficients have potentially wide ranges (minimum and maximum values difference) with many in-between values that depend on the increment step size (Inc). The so called *Exhaustive coefficient variation* (Table 6) can be used to systematically investigate coefficient ranges in search for coefficient values that provide the best statistical fit. For example CSDL coefficient value for minimum 10.0, maximum 14.0 and increment step 0.2 are passed into the cultivar file and model is executed. The same procedure is repeated with SLVR coefficient with minimum 200, maximum 400 and increment step 10. In these simple examples as shown in Table 6 it can be seen that for a coefficient range from 10.0 to 14.0 with increment steps of 0.2 a total 21 coefficient variations are executed for the CSDL coefficient. For SLVR, cultivar coefficient with a range from 200 to 400 with an increment step of 10 a total of 21 coefficient variations are executed. In order to overcome time losses in the process of cultivar coefficients estimation based on the statistical fit, a range reduction method (Röll et al. 2020) was implemented (Table 6). With range reduction method three global rounds are conducted in the process of estimating cultivar coefficient with smallest average nRMSE. Greater increment steps are used in the first round for each given coefficient range with CSDL coefficient having minimum 10.0, maximum 14.0 and increment step 1.0 and SLVR having minimum 200, maximum 400 and increment step 50. Based on the lowest AVG-nRMSE, the value for each coefficient value is selected, i.e. 12.0 for CSDL and 300.0 for SLVR. In the second round, the new coefficient ranges with a narrower increment step are executed with CSDL having minimum 11.6, maximum 12.4 and increment step 0.4 and SLVR having minimum 280, maximum 320 and increment step 20. Based on the lowest AVG-nRMSE, the new coefficient "optimums" are selected, i.e. (CSDL=12.4, SLVR=320). In the final round new ranges for each coefficient are defined with CSDL having minimum 12.2, maximum 12.6 and increment step 0.2 and SLVR having minimum 310, maximum 330 and increment step 10. The final values are selected based on lowest AVG-nRMSE (CSDL=12.2, SLVR=310). Based on the range reduction approach, 48% fewer combinations are executed when compared to exhaustive coefficient variations. The range reduction method as described in Table 6 is expected to retain a systematic optimum localisation approach (achievable with exhaustive gridding variation), and provides more realistic values for the coefficients when compared to the random generation of cultivar coefficients for allowed ranges.

Table 6 Comparison of examples of exhaustive coefficient variation and range reduction method with total number of coefficient variations (cv. 'Bragg' *DSSAT Default* cultivar coefficients, Gainesville 1978)

Exhaustive variation		**Range reduction method**					
CSDL							
Minimum	10.0	10.0		11.6		12.2	
Maximum	14.0	14.0		12.4		12.6	
Increment step	**+0.2**	**+ 1.0**		**± 0.4**		**± 0.2**	
	10.0	10.0					
	10.2	11.0		11.6			
	…	12.0	**→12.0**	12.0		12.2	**→12.2**
	→12.2	13.0		12.4	**→12.4**	12.4	
	…	14.0				12.6	
	13.8						
	14.0						
Number of combinations:	21	5		3		3	
Total	**21**			**11**			
SLVR							
Minimum	200.0	200.0		280.0		310.0	
Maximum	400.0	400.0		320.0		330.0	
Increment step	**+10.0**	**+ 50.0**		**± 20.0**		**± 10.0**	
	200.0	200.0					
	210.0	250.0		280.0			
	…	300.0	**→300.0**	300.0		310.0	**→310.0**
	→310.0	350.0		320.0	**→320.0**	320.0	
	…	400.0				330.0	
	390.0						
	400.0						
Number of combinations:	21	5		3		3	
Total	**21**			**11**			

- *CSDL* and *SLVR* are defined in Table 3