User guidelines for running TSE software
(Time-Series cultivar coefficient Estimator for DSSAT crop models)
Standalone external DSSAT plug-in

Emir Memic (University of Hohenheim, Stuttgart, Germany, April 2023)
E-mail: emir.memic@uni-hohenheim.de; emir_memic@windowslive.com

TSE **T**ime-**S**eries cultivar coefficient **E**stimator

This cultivar coefficients optimization program was developed and tested with DSSAT4.7 (first TSE release) CROPGRO-Soybean model. The program was partially tested with CERES-Maize and Wheat models. The program was designed and written to enable optimization of cultivar and ecotype coefficients of all available crop models in DSSAT4.7 and future DSSAT4.8 release and should work with other future versions. For writing this user guidelines DSSAT 4.8 version was used.

TSE_v1 – (June_2021)

TSE_v2.2 (Latest official release of TSE version 2) – (April_2023)

New features in version 2:

- Full window view will expand interface to fit the screen.

- Default resolution:
  1410x760

- Scale layout:
  - The computer screen scale layout must be set to 100% for fully operable TSE!

  - Scale layout set at 125% is displayed properly to a certain degree (TSE partially operable).

  - Windows scale layout 150% will not display TSE options properly! (TSE inoperable!)

- PlantGro.OUT default DSSAT time-series output file used! Additionally other DSSAT time-series output files can be used such as: PlantGr2.OUT etc.
- Multiple @TRNO layers in File-T can be used directly in the optimization process.
- Boxplot sensitivity analysis output available in TSE_workspace for selected target variables based on used coefficients in the optimization process.

- Debugging:
  - captureError.exe - Additional debugging option with “captureError.exe” in TSE_v2.2 directory (**C:\DSSAT48\Tools\TSE_v2.2).** After captureError.exe is executed a short description of the tool is explained**.**

Contents

I The TSE concept overview (Memic et al. 2021)
The cultivar/ecotype 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 (File-X), weather file, soil profile, functional species, ecotype and cultivar files, time-series file containing in-season observations (File-T) and summary file (File-A - optional) are available.

Overall program run can be separated in three steps:

1. Selection of the File-X and corresponding treatments defined in File-X. Based on the selected experiment corresponding in-season observations (as time-series data) are read in as temporal inputs (all available time-series observations can be used), and cultivar coefficients designated for optimization (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 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".

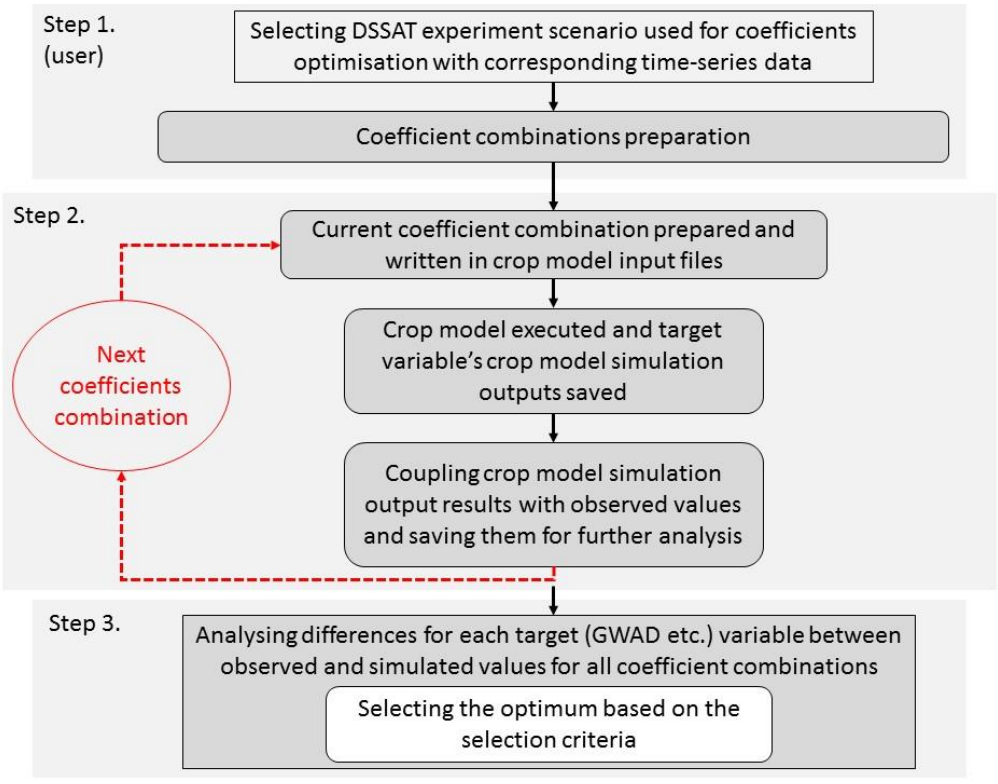


Figure 1: Flow chart showing overall program run in three steps (GWAD – grain weight) (Memic et al. 2021)

The program can be used for optimizing cultivar coefficients based on single or multiple experiment data sets. User should first optimize phenology-related cultivar coefficients. After phenological events are correctly optimized, growth-related cultivar coefficients are optimized. Phenology-related cultivar coefficient optimization is not conducted based on the time-series data (File-T), but by using File-A observations and by simply minimizing the difference between simulated and observed phenological event as day after planting. Growth-related cultivar coefficients are optimized by using File-T in-season observations by means of normalized RMSE (among others) 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.

II General TSE program settings overview

The program was developed (in Windows environment, for use on Windows) as standalone external DSSAT plug-in and has no specific installation requirements. All required files for running the program are shared on GitHub via zipped file.

The "**TSE_v2.2.zip**" file must be unzipped and copied to the Tools directory: "**C:\DSSAT48\Tools**" (Figure 2) (depending on the version "C:\DSSAT**\Tools").

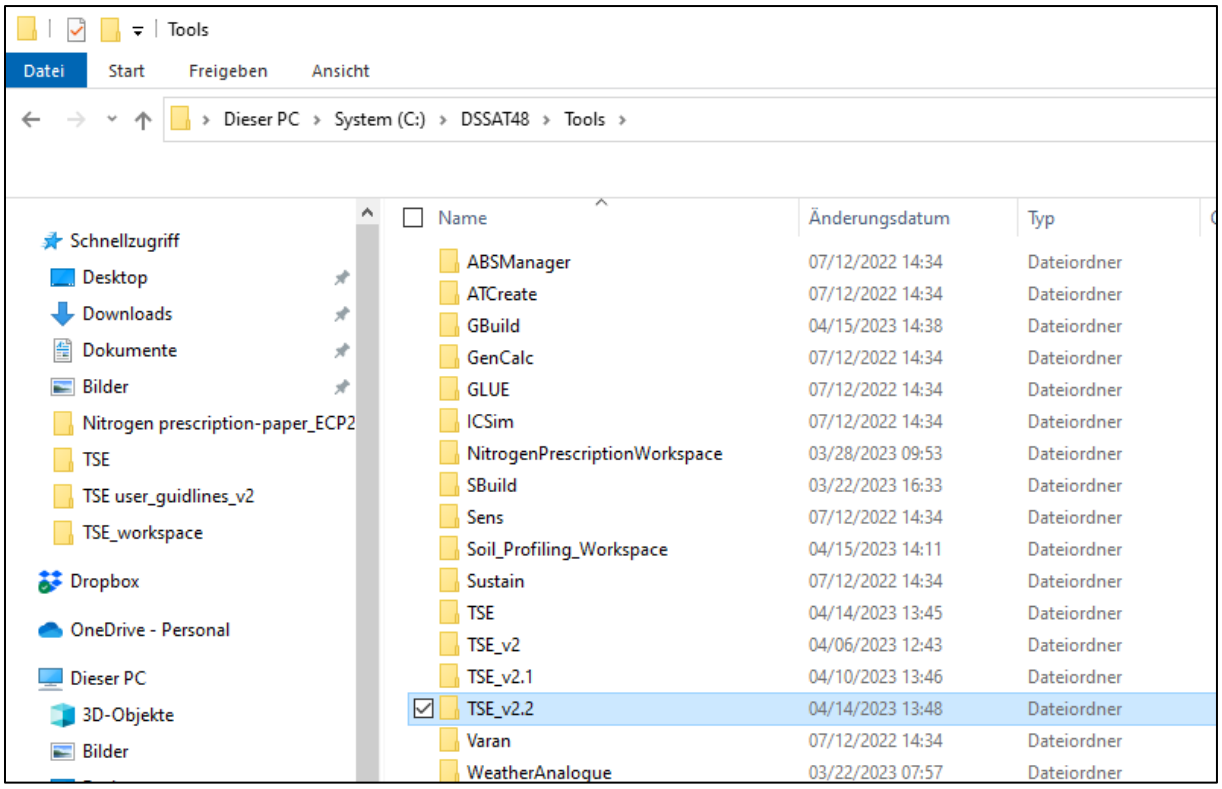


Figure 2 "C:\DSSAT48\Tools\TSE_v2.2"

In the folder "TSE_v2" **"C:\DSSAT48\Tools\TSE_v2.2"** (Figure 3) **"TSE_v2.2.exe"** windows runnable must be **executed as "<u>Administrator</u>"** (Figure 3).

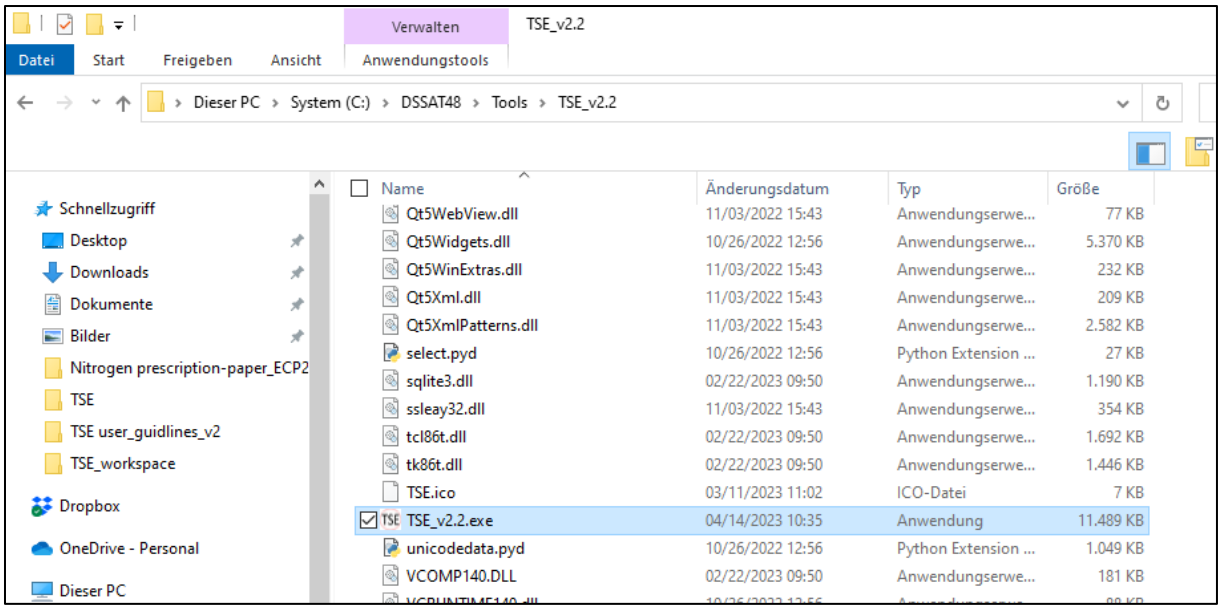


Figure 3 **"TSE_v2.2.exe"** windows runnable

VERY IMPORTANT:

I. **PlantGro.Out** (or other time-series) crop model outputs are coupled to those in **File-T** (phenology- or growth-related) time-series in-season observations.

II. **Evaluate.Out** crop model outputs are used for optimising Phenological events which are passed into the model through **File-A**.

III. If sub-model (e.g. WHAPS) is initialised in the File-X, the calibrator might NOT work! (in File-X in *SIMULATION CONTROLS in GENERAL line, column SMODEL **do NOT initialize sub-models such as WHAPS, IXIM etc.!**). This does TSE program.

IV. For multi-TRT optimizations only target variables simultaneously available in all **File-T**/s (for corresponding **File-X**/s Treatment/s) are accessible for optimisation.

V. **File-T** observations: **All in-season observations available including 0 are used!** *Only "-99" values are ignored by TSE*.

VI. The program is matching DOY from **File-T** with those in the **PlantGro.OUT**. If the user setup in File-X reporting frequency for example every fifth day and exact observation DOY is not present in the PlantGro.OUT as it is written in the File-T, the program will NOT be able to match them for comparing simulated with observed.

The TSE program is creating additional directory "**TSE_workspace**" ("**C:\DSSAT48\TSE_workspace**") (Figure 4) and modifying the cultivar file in that directory, which is then executed by main DSSAT model executable.

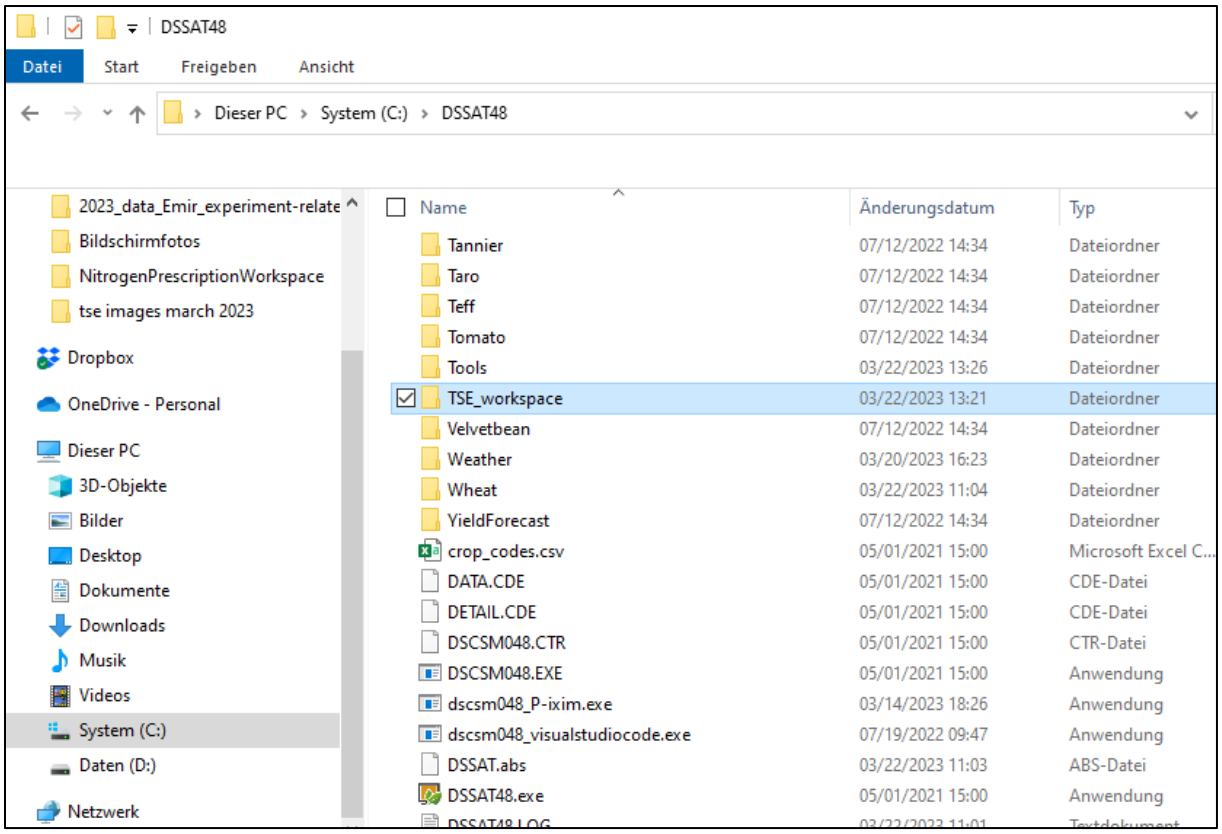


Figure 4 ("**C:\DSSAT48\TSE_workspace**")

The TSE program does NOT modify core DSSAT files in their original directories!

Program run is considered: From when **"TSE_v2.2.exe"** is **executed** until "**Exit**" push button is pressed. Any form of optimization done in-between is temporarily saved in the cultivar file in "TSE_workspace" directory as a working version.

After TSE program is started (**"TSE_v2.2.exe"** executed) all modifications on Cultivar or Ecotype file are conducted in ("**C:\DSSAT48\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:\DSSAT48\TSE_workspace**. If user is satisfied with the cultivar coefficient values based on nRMSE (or other) fit, cultivar coefficient combination MUST be copied to "**C:\DSSAT47\Tools\Genotype**" located cultivar or ecotype file **MANUALLY**. If TSE program is started again without saving the combination in **C:\DSSAT47\Tools\Genotype** located cultivar or ecotype file new TSE program start will copy original cultivar or ecotype file located in C:\DSSAT47\Tools\Genotype and overwrite user's working cultivar file in "**C:\DSSAT48\TSE_workspace**".

After model run finished and before the user clicks "**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:\DSSAT48\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 (time constraint)**.

Example of optimizing cultivar coefficient (*CROPGRO-Soybean, SBGRO048.CUL*):
Every time **TSE_v2.2.exe** is executed, original cultivar file (SBGRO048.CUL) file from **C:\DSSAT48\Tools\Genotype** will be copied to "**C:\DSSAT48\TSE_workspace** directory (Figure 5) 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\Tools\Genotype** directory into SBGRO048.CUL **Manually**.

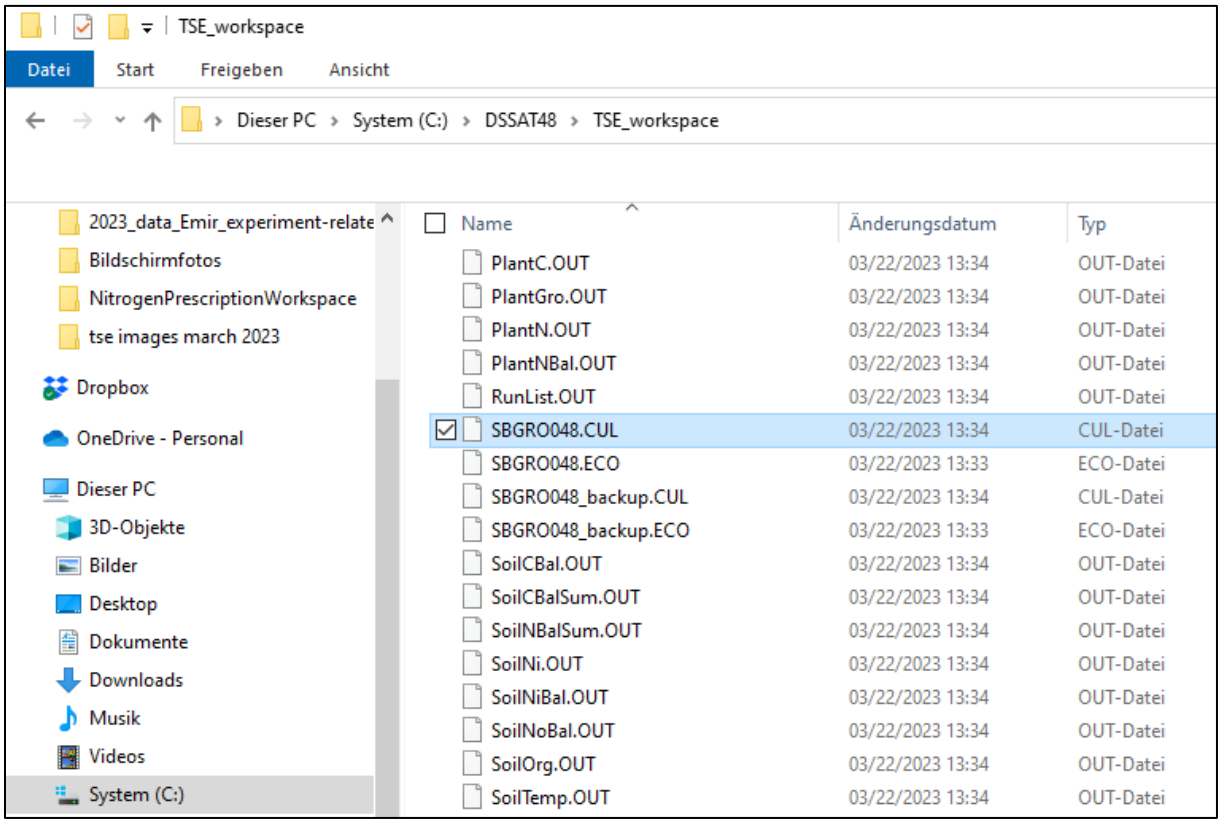


Figure 5 C:\DSSAT48\TSE_workspace\SBGRO047.CUL

The original cultivar is saved as **"!Old_timestamp_cultivarID…."** (Figure 6, text editor line 75-76) and the new one (Figure 7, text editor line 73) is saved in that working cultivar file (in **TSE_workspace**) and model is executed.

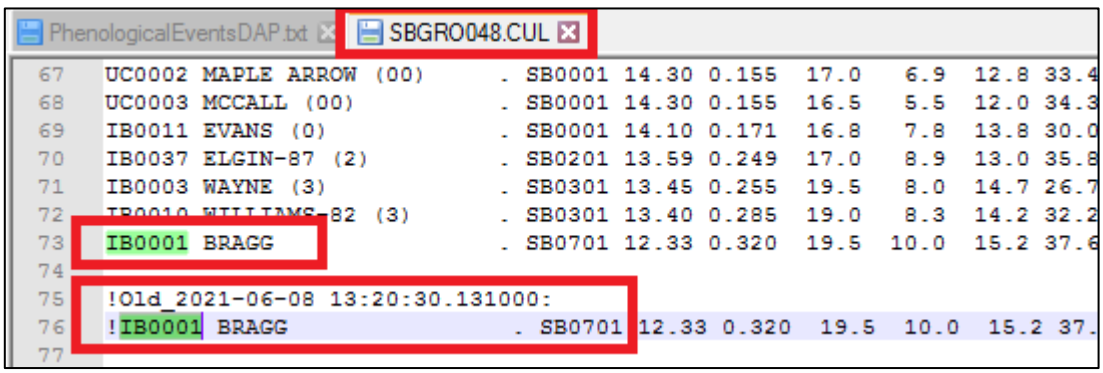


Figure 6 Old and new cultivar

III Running TSE program

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

1. **If directory path shown is "C:\DSSAT48" (**or other C:\DSSAT****), do NOT modify! If the path is not "C:\DSSAT48" (This means that TSE folder was not copied to the "C:\DSSAT48\Tools"), then use "Change Dir." To navigate to "C:\DSSAT48" folder and select it. This will be explained later in more details.**

2. **Select desired model, cultivar or ecotype and Initialize selected!**

3. **Select cultivar from model corresponding list and File/s-X from list containing selected cultivar.**

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

5. **Execute selected treatment/s with DSSAT model to check if core DSSAT files are runnable and select Default/Advanced.**

6. **Select optimization of Phenology/Growth -related coefficients and corresponding methods.**

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

8. **Check if optimization software setup is correct.**

9. **Run the model! ("*Run – Range setup"* or *"Run – Range reduction method"*)**

10. **Estimate Multi-treatment-based cultivar coefficient combination or execute GBuild!**

Interface

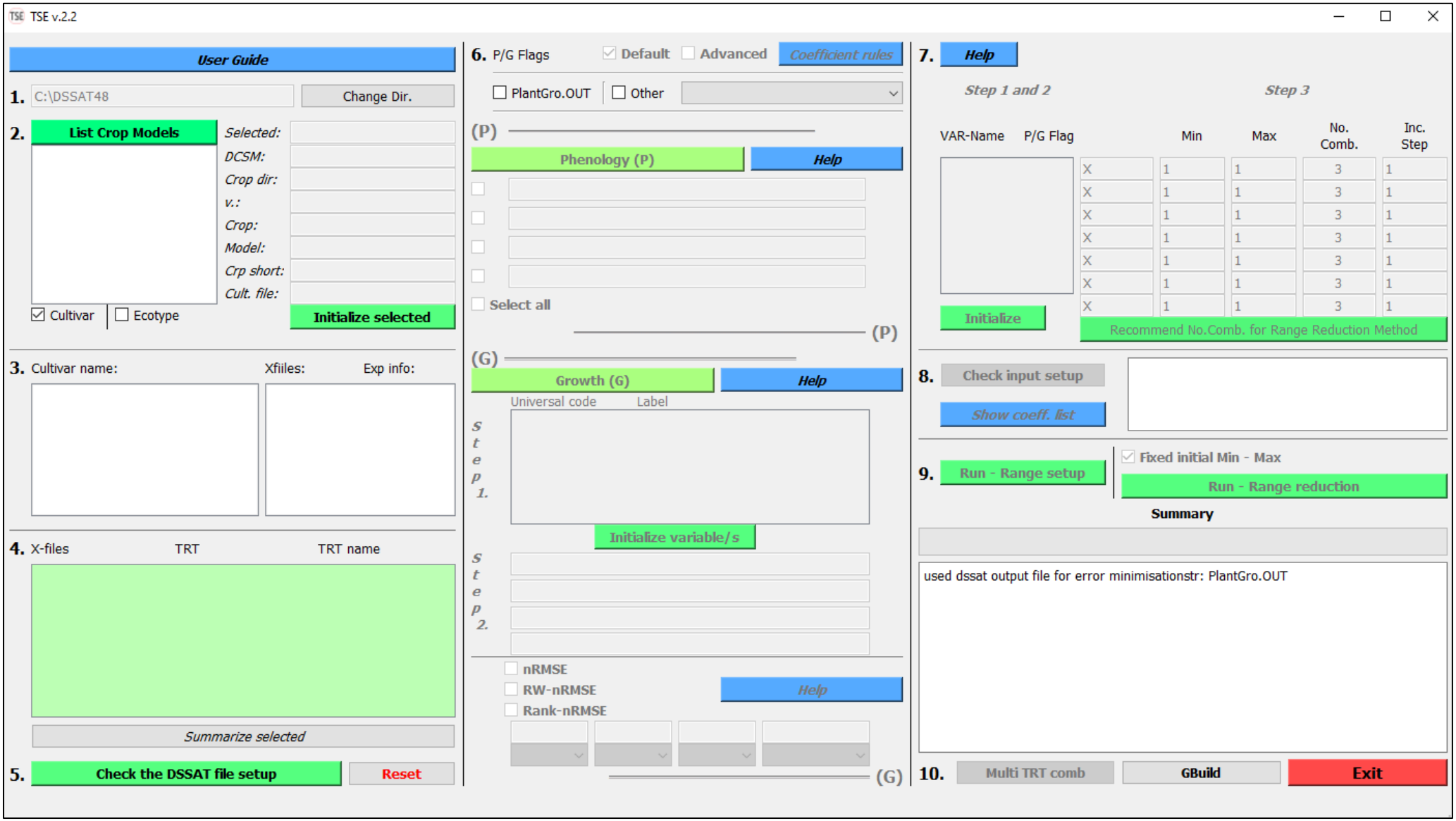


Figure 7. Interface

1. Do NOT modify!

If directory path shown is "C:\DSSAT48" (depending on what version of the model user is running), do ***NOT*** modify! If the path is not "C:\DSSAT47 or C:\DSSAT48" or other, then copy TSE_v2 folder into the "C:\DSSAT**\Tools" (Figure 8). After re-executing **"TSE_v2.2.exe"** from "C:\DSSAT48\Tools\TSE_v2.2" the correct path should be read in.

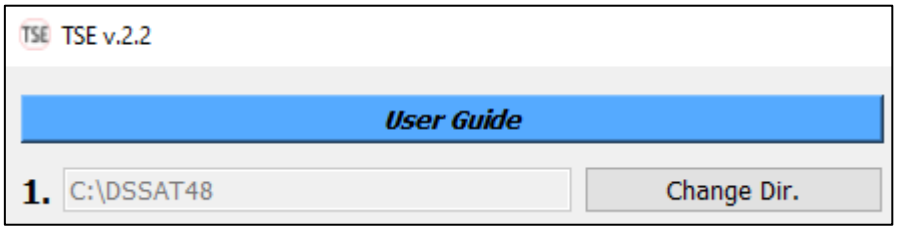


Figure 8. DSSAT directory path

2. Select desired model, cultivar or ecotype and Initialize selected!

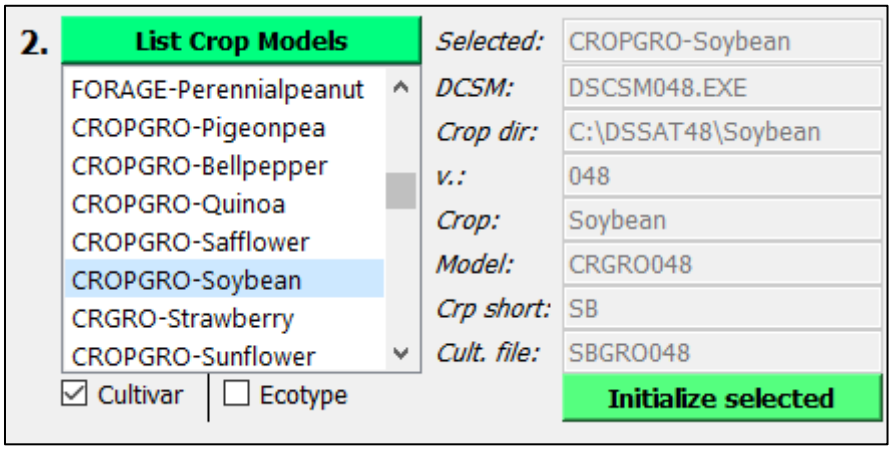


Figure 9. The list of models available in the DSSAT are uploaded into the TSE interface after List Crop Model is clicked. After selecting desired model and cultivar or ecotype the Initialize selected push button has to be clicked.

3. Select cultivar from model corresponding list and File/s-X from list containing selected cultivar.

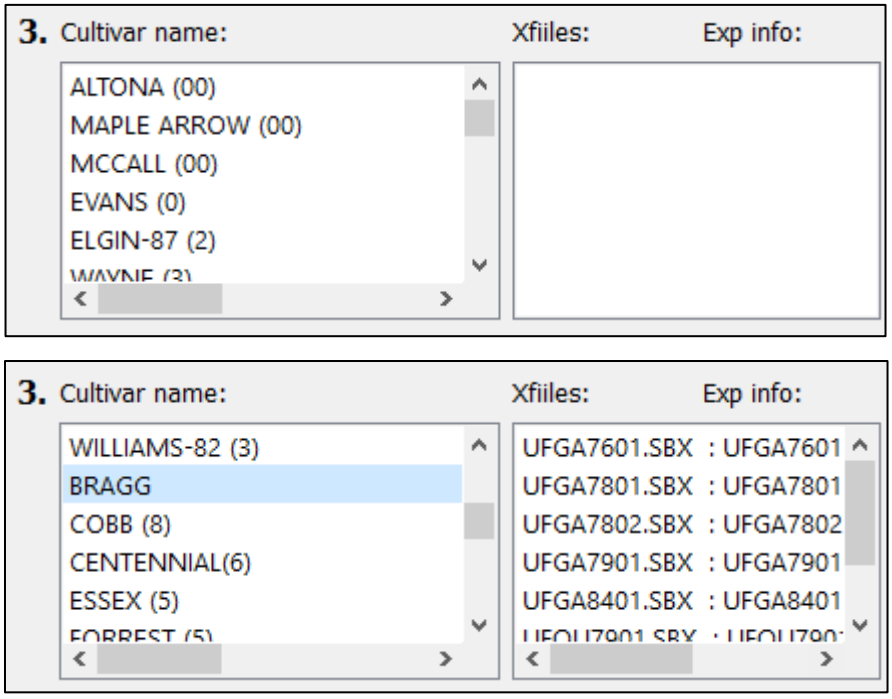


Figure 10a: After the model is initialized the list of the model corresponding cultivar names are read in the interface ("Cultivar name"). After cultivar name (e.g., "BRAGG") is selected the File/s-X containing that cultivar name are read in the interface ("X files and Exp info") (Figure 10b).

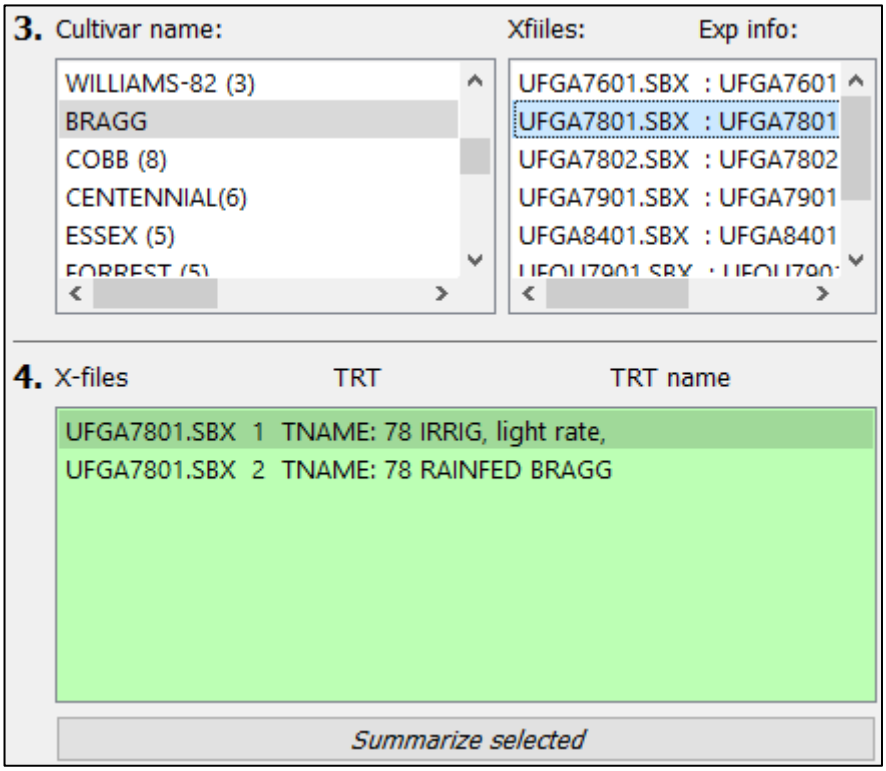


Figure 11: In the list File-X (or multiple File-Xs) can be selected as showed (e.g., UFGA7801.SBX) (Multiple treatment (experiment) selection is done by: ***Ctrl+ mouse left button click***).

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

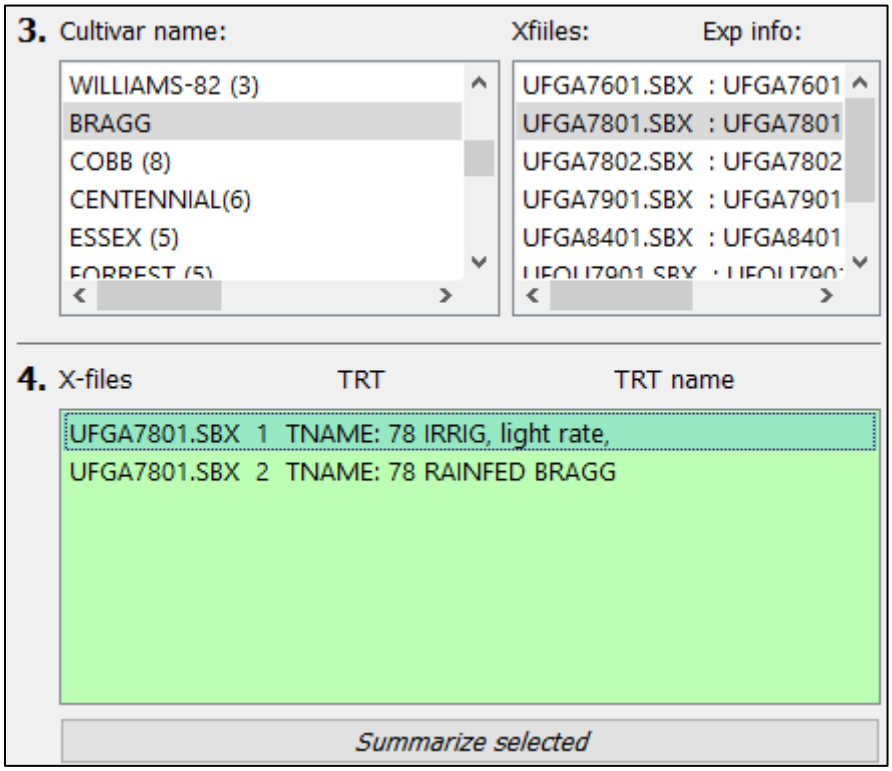


Figure 12: After selecting File-X in the interface the list of the treatments in the File-X containing the cultivar is shown and offered for selection (green list widget box). The green box (green list widget box in the interface) is where the experiment file and treatments for optimizing cultivar coefficients are **selected**. Multiple treatments can be selected (multiple treatment (experiment) selection is done by: ***Ctrl+ mouse left button click***).

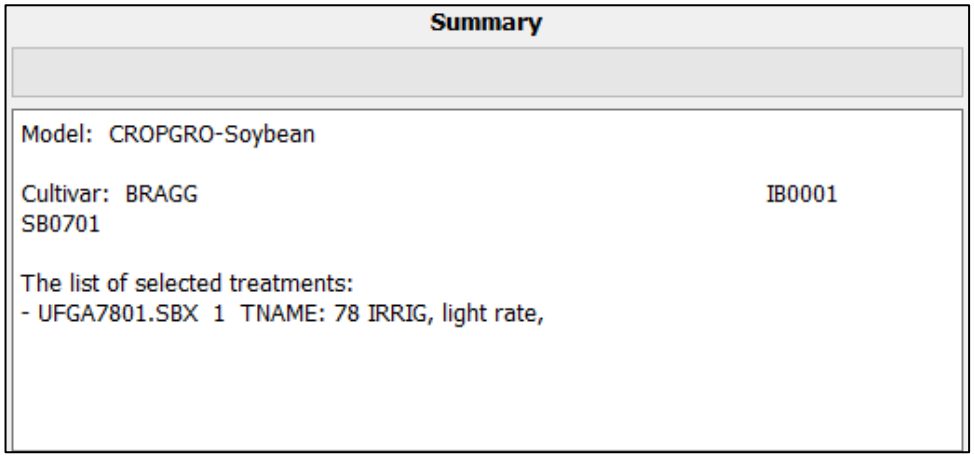


Figure 13: The "Summarize selected" push button (Figure 12) will display selected options in the Summary window in the interface.

5. Execute (Check the DSSAT file setup) selected treatment/s with DSSAT model to check if core DSSAT files are runnable.

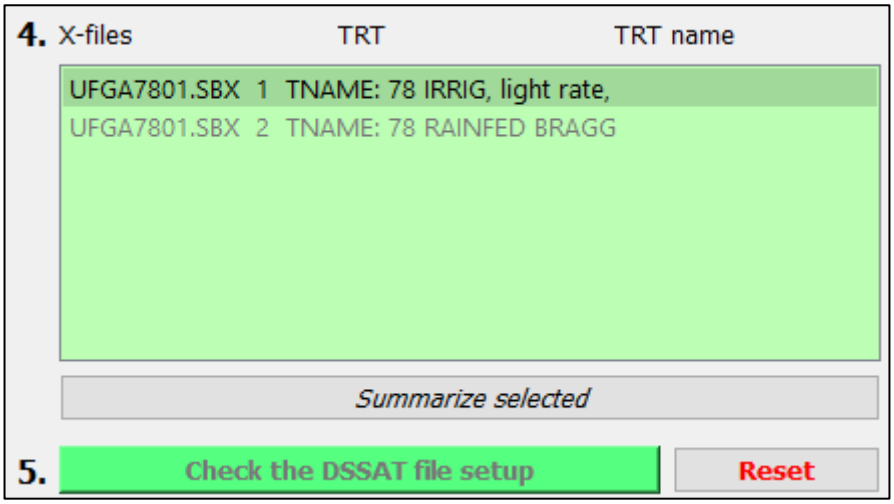
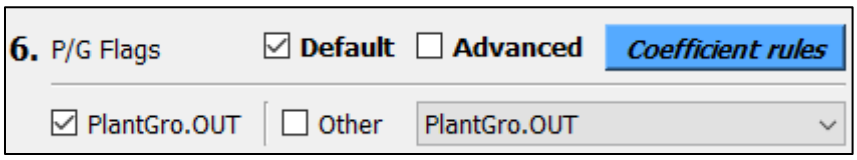


Figure 14: The "**Reset**" button will unlock the access to the selection of the crop model, cultivar/ecotype and File-Xs.

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

Cultivar specific coefficient flags (Phenology and Growth – P/G) are based on cultivar file (in-file "!Calibration" line starting with DSSAT4.8 version or are listed in the "C:\DSSAT48\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) (Figure 15). ***Advanced*** (check box) will upload all available cultivar coefficients from cultivar file for potential optimization (Figure 15). It is possible to use other DSSAT time-series output files for coefficient optimization in case observations are simultaneously available in corresponding output file (e.g., PlantGr2.OUT) and File-T (Figure 15).

Phenology:

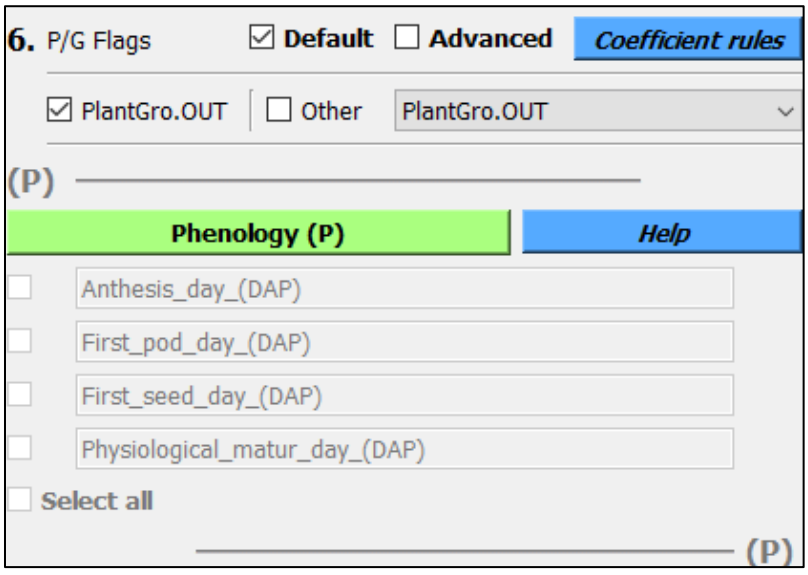


Figure 16: There are four phenological events parameters described in (Memic at al. 2021) hardcoded in the TSE algorithm, specifically for CROPGRO-Soybean.

This hard coded solution can be over-written with manually customizable text file located in "C:\DSSAT48\Tools\TSE" named

"PhenologicalEventsDAP.txt" (Figure 17). The file setup is explained in the file as shown in following figure and has to be located in "C:\DSSAT47\Tools\TSE" directory and format rules have to be abided!

PhenologicalEventsDAP.txt

```
1  ! Phenological events as Days after planting (DAP)
2  !
3  !  Only four can be read in and the rest should be outcomented with "!"
4  !
5  !  Phenological events can be added to this list!
6  !  In order to work Phenological event - code (etc. ADAP) has to be outputed
7  !  by DSSAT in the Evaluate.OUT and this format has to be followed:
8  !  Example:
9  !  Code      Code_Name
10 !  ADAP      Anthesis_day
11 !
12 !---------------------------------------!
13 ADAP         Anthesis_day_(DAP)
14 PD1P         First_pod_day_(DAP)
15 PDFP         First_seed_day_(DAP)
16 MDAP         Physiological_matur_day_(DAP)
17 !PD1T        First_pod_date_(DAP)(YrDoy)_?
18 !EDAP        Emergence_date_(DAP)
19 !B1DAP       Branch_1_date_(DAP)
20 !B2DAP       Branch_2_date_(DAP)
21 !B3DAP       Branch_3_date_(DAP)
22 !B4DAP       Branch_4_date_(DAP)
23 !FDAP        Forcing_Date_(DAP)
24 !TDAP        Begin_tuber_growth_(DAP)
25 !---------------------------------------!
```

Figure 17. PhenologicalEventsDAP.txt

Growth:

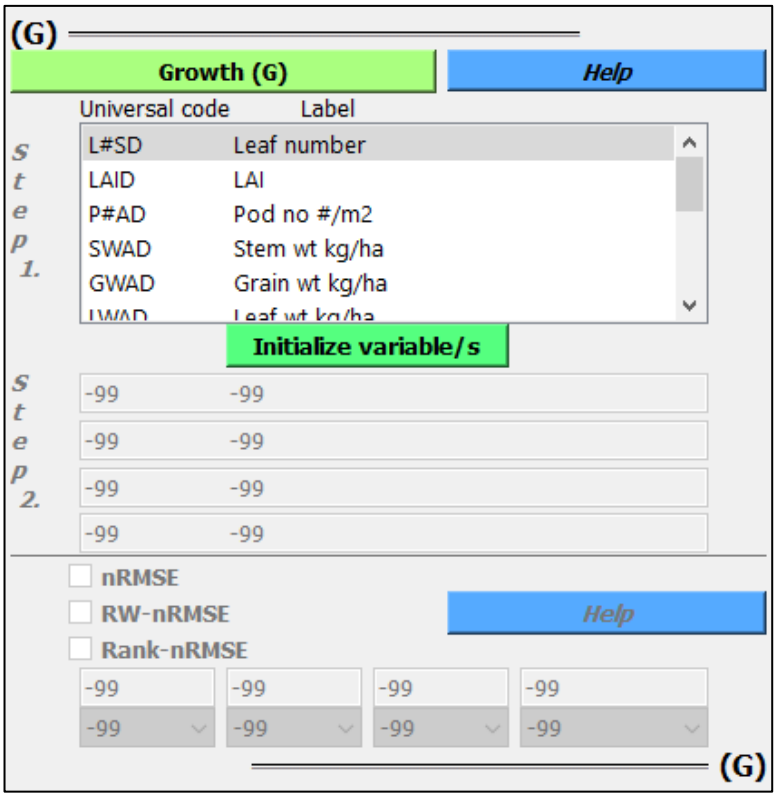


Figure 18: In this step, error minimization method (nRMSE or RW-nRMSE or Rank - nRMSE, explained in the APPENDIX) is selected. Maximum four different target variables (e.g. LWAD, CWAD …) can be selected. If four are not available, one or two or three can be selected. In the first round of optimization four can be selected and in the second four, or with some other combinations.

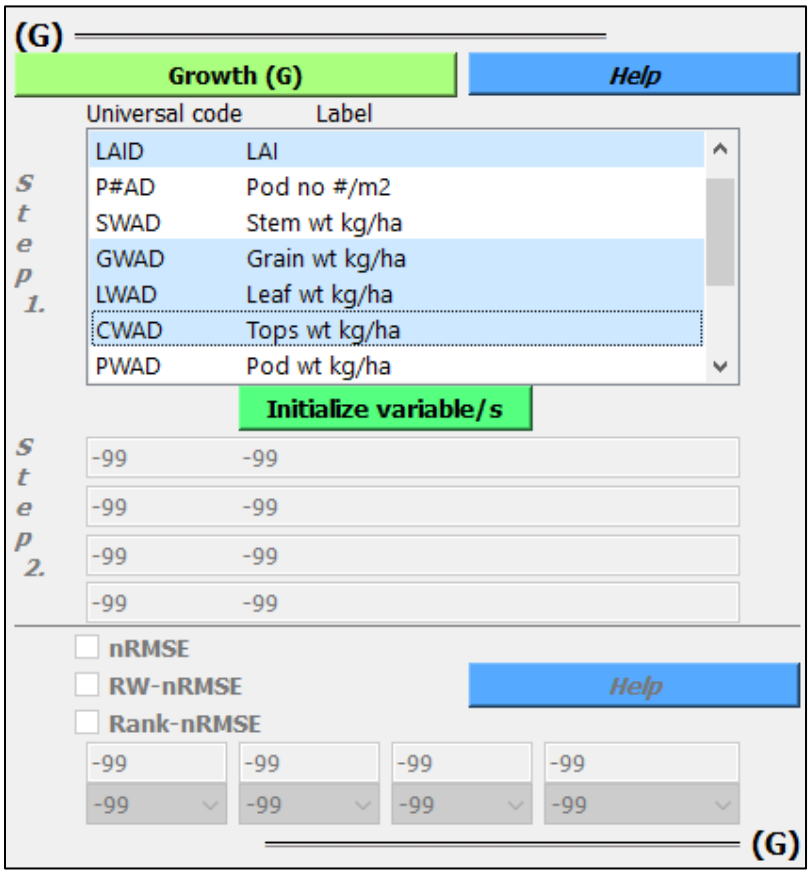


Figure 19: "***Ctrl* + *mouse left click***"- for selecting multiple target variables simultaneously!

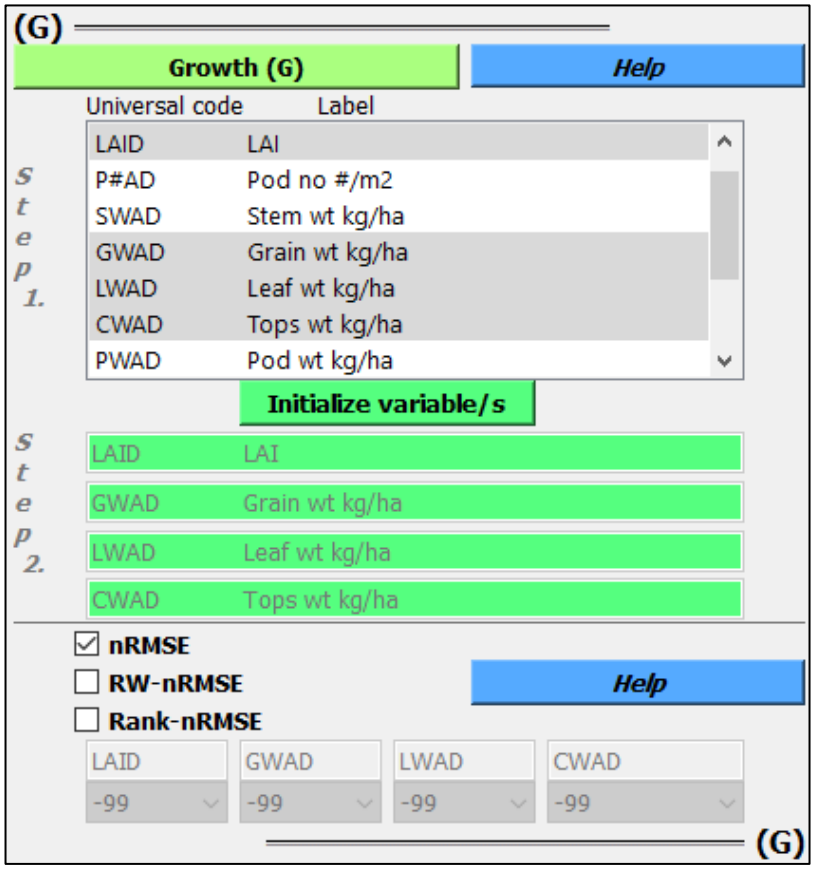


Figure 20: After selecting target variables Initialize target variable/s has to be clicked.

7. <u>Selecting</u> desired coefficients and coefficient ranges and increment steps.

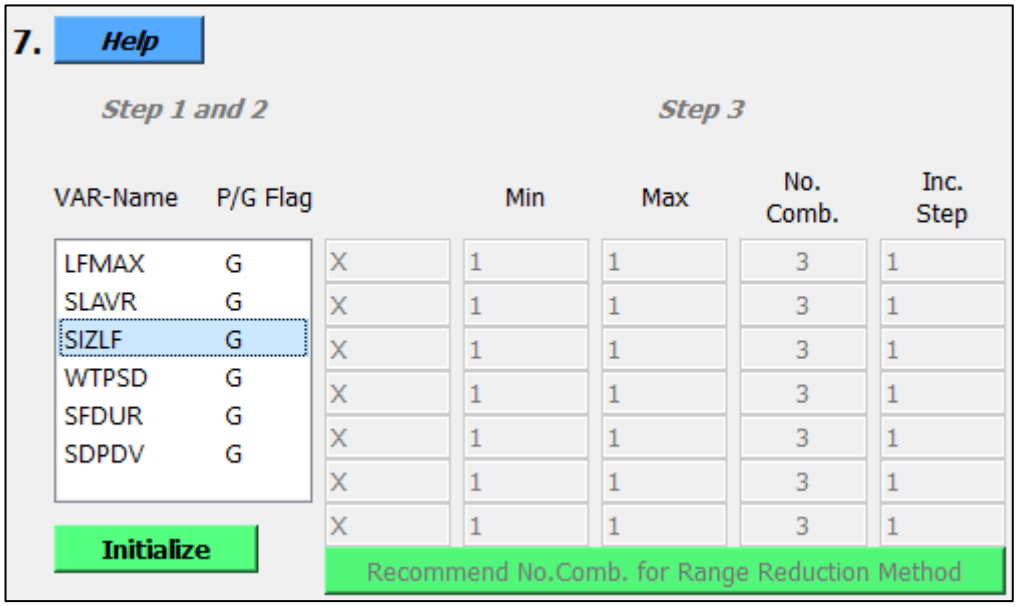


Figure 21: First: coefficient/s are selected from the list:

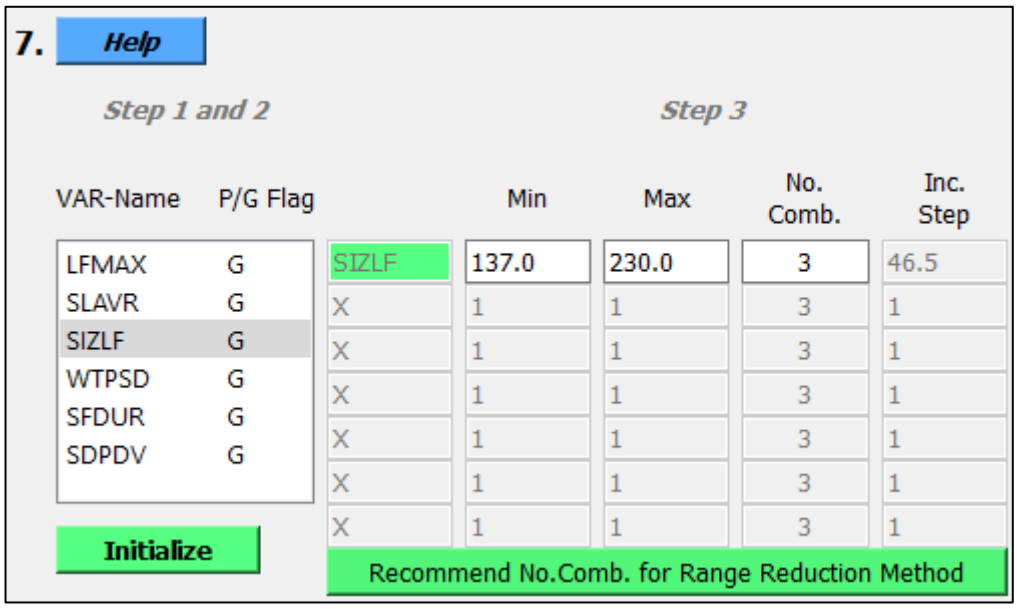


Figure 22: After selecting coefficient or multiple coefficients (***Ctrl+ mouse left button click*** – for multiple selection) "Initialize" push button is clicked!

<u>In the DSSAT4.8version additional MINIMA/MAXIMA coefficient values are included in cultivar files with more details. Since MINIMA/MAXIMA are recommendations made by crop model developers a new code addition was done to the TSE algorithm. The TSE program will check for these values and populate Min Max edit boxes with those values (Figure 22).</u>

Initialize push button 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 2.) (Figure 22). Automatically this will create a simple range of Min/Max values with three coefficient combinations that are going to be passed into the cultivar file, after each model is executed (Figure 22). If the user wants more combinations in between given Min/Max range number of desired coefficient combinations between Min/Max can be given in "No. Comb." edit boxes (Figure 22). After giving the desired number of combinations between Min/Max increment steps are automatically calculated (Figure 22). New option

Recommend No. Comb for Range Reduction Method will automatically suggest increment step with No. Comb. If too many recommended that might result in too long optimization session, the user can reduce them.

If a user would like to modify Min/Max values, it can be done directly (manually) in edit boxes below Min/Max labels in step 2 (Figure 22). If Min/Max are modified based on the number of combinations values are automatically calculated and populated in corresponding edit boxes.

If the user wants to fix the value of some coefficient, then they set Min and Max to equal value (same value in the Min edit box as in the Max value edit box).

8. Check if optimisation software setup is correct.

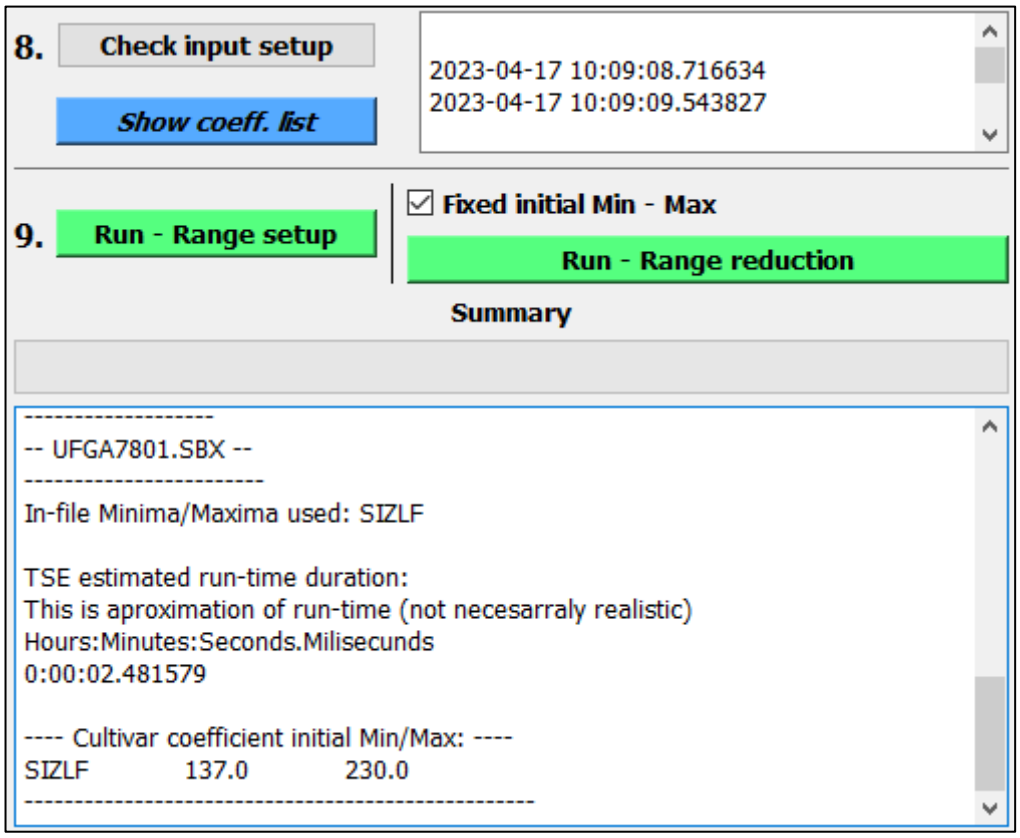


Figure 23: Show coeff. List push button (section 8.) will display coefficient to be used in the estimation process (Figure 24).

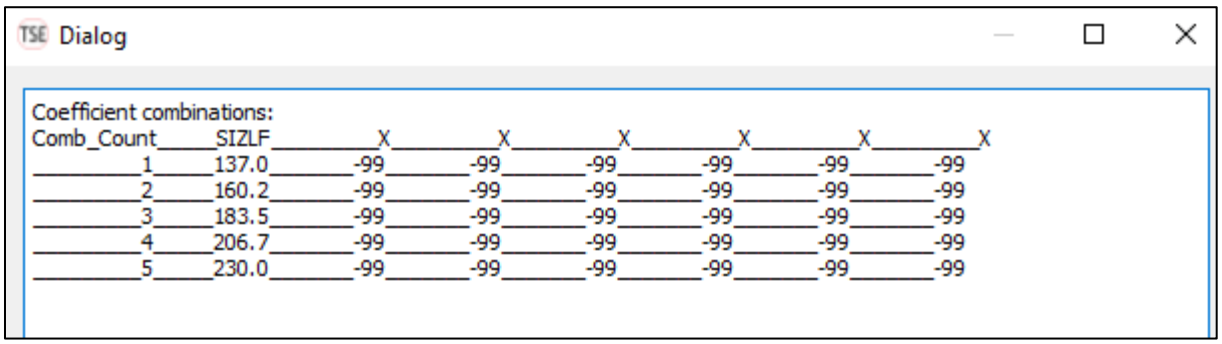


Figure 24. Summery list of coefficients to be used in optimization process.

9. ***Run – Range setup*** *&* ***Run – Range reduction****!*

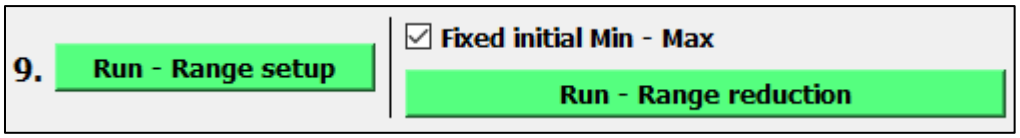
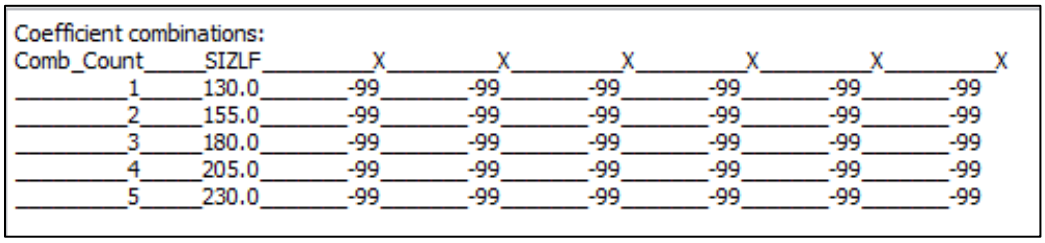


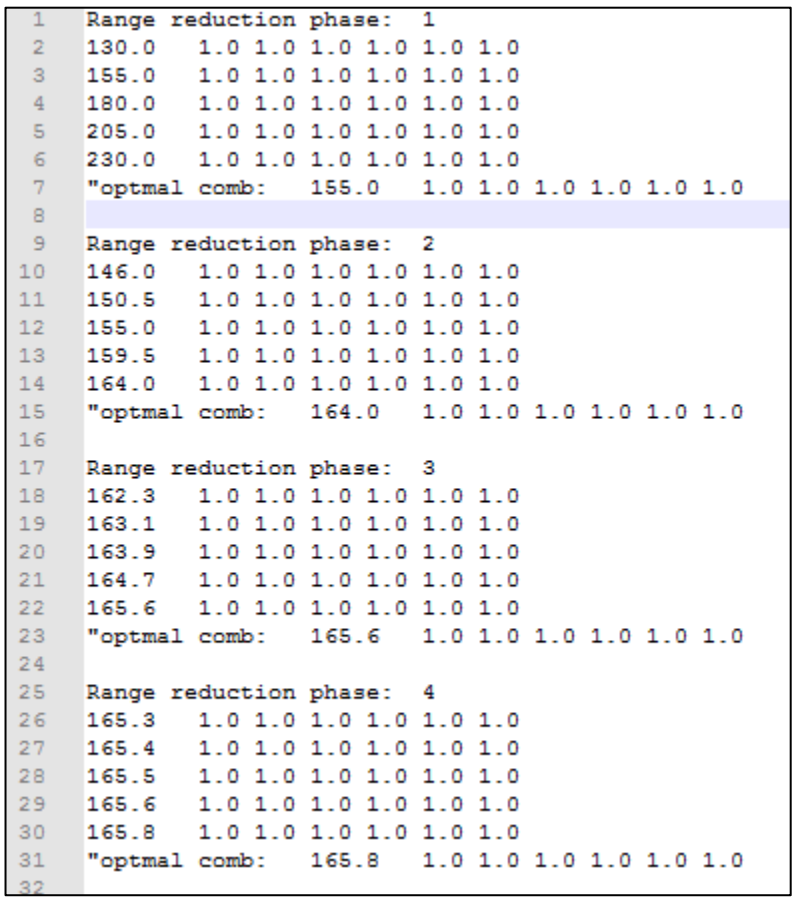
Figure 25: Selection of the error minimization methods. Range reduction is described as an example in Appendix A.5.



Coefficient combinations:

Comb_Count	SIZLF	X	X	X	X	X	X
1	130.0	-99	-99	-99	-99	-99	-99
2	155.0	-99	-99	-99	-99	-99	-99
3	180.0	-99	-99	-99	-99	-99	-99
4	205.0	-99	-99	-99	-99	-99	-99
5	230.0	-99	-99	-99	-99	-99	-99

Figure 26: **Run-Range setup** – will pass to the cultivar/ecotype file only these five coefficients based on the example described in user guidelines section *7*.



```
Range reduction phase:  1
130.0   1.0 1.0 1.0 1.0 1.0 1.0
155.0   1.0 1.0 1.0 1.0 1.0 1.0
180.0   1.0 1.0 1.0 1.0 1.0 1.0
205.0   1.0 1.0 1.0 1.0 1.0 1.0
230.0   1.0 1.0 1.0 1.0 1.0 1.0
"optmal comb:   155.0   1.0 1.0 1.0 1.0 1.0 1.0

Range reduction phase:  2
146.0   1.0 1.0 1.0 1.0 1.0 1.0
150.5   1.0 1.0 1.0 1.0 1.0 1.0
155.0   1.0 1.0 1.0 1.0 1.0 1.0
159.5   1.0 1.0 1.0 1.0 1.0 1.0
164.0   1.0 1.0 1.0 1.0 1.0 1.0
"optmal comb:   164.0   1.0 1.0 1.0 1.0 1.0 1.0

Range reduction phase:  3
162.3   1.0 1.0 1.0 1.0 1.0 1.0
163.1   1.0 1.0 1.0 1.0 1.0 1.0
163.9   1.0 1.0 1.0 1.0 1.0 1.0
164.7   1.0 1.0 1.0 1.0 1.0 1.0
165.6   1.0 1.0 1.0 1.0 1.0 1.0
"optmal comb:   165.6   1.0 1.0 1.0 1.0 1.0 1.0

Range reduction phase:  4
165.3   1.0 1.0 1.0 1.0 1.0 1.0
165.4   1.0 1.0 1.0 1.0 1.0 1.0
165.5   1.0 1.0 1.0 1.0 1.0 1.0
165.6   1.0 1.0 1.0 1.0 1.0 1.0
165.8   1.0 1.0 1.0 1.0 1.0 1.0
"optmal comb:   165.8   1.0 1.0 1.0 1.0 1.0 1.0
```

Figure 27: **Run-Range reduction** – will use initial Min/Max and increment step as a starting point and through four run cycles try to localize (find local optimum) the best coefficient value based on the error minimization method for coefficient values that are not tested at all in simple "Run-Range setup", as shown in following figure, as described in APENDIX 5.

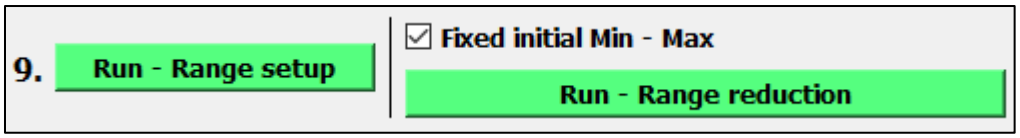


Figure 28: If "Fixed initial Min – Max" check box is "checked" range reduction algorithm will not be allowed to offer Min "optimum" lower than that which is defined while initializing cultivar coefficient values. Same goes for Max values.

10. Additional options

Multi TRT comb

If this push button clicked multi treatment-based cultivar coefficient combination will be estimated, as described in appendix A.2 (Figure 29).

Gbuild

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

If the **PlantGro.Out** in "**C:\DSSAT48\TSE_workspace**" is open with GBuild visual and statistical fit across all used experiments of coefficient optimization can be seen in GBuild (Figure 29).

Exit the program and all running threads.

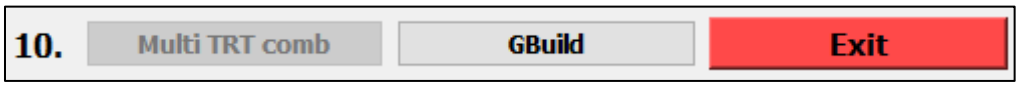


Figure 29: Multi TRT comb, Gbuild and Exit

IV Boxplot based coefficient sensitivity analysis

Boxplots generated by TSE algorithm can be used as indicator of sensitivity of specific target variable to different coefficient values.

It must be pointed out that cultivar coefficient estimation process is not purely mathematical or statistical exercise. Cultivar coefficients must be based on certain agronomic criteria. Just because some coefficient combinations provide good statistical fit, at cost of real morphological criteria, it does not mean that those should be used. For example: if coefficient defining leaf size is not realistic, but provides better statistical fit for the data, IT SHOULD NOT BE USED!

Keeping this in mind generated boxplots can be used to investigate to what degree TSE defined Min/Max/Inc values for specific coefficient have influence on measured target variables. One example of using boxplots is given in Figure 30.

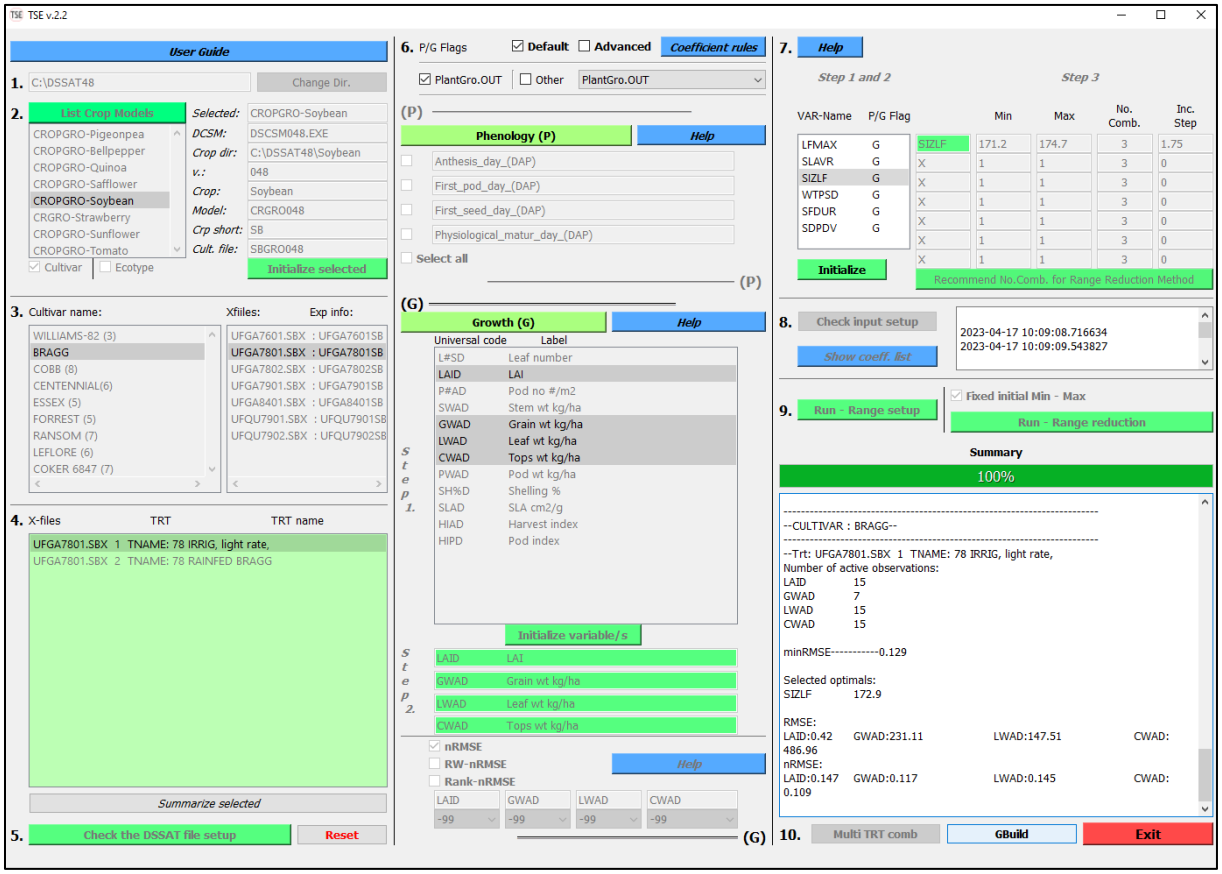


Figure 30: This TSE setup with SIZLF coefficient when executed with “Run-Range Setup” and “Run-Range reduction”

1. *Range setup example*

C:\DSSAT48\TSE_workspace -> UFGA7801.SBX_1_boxPlot.png

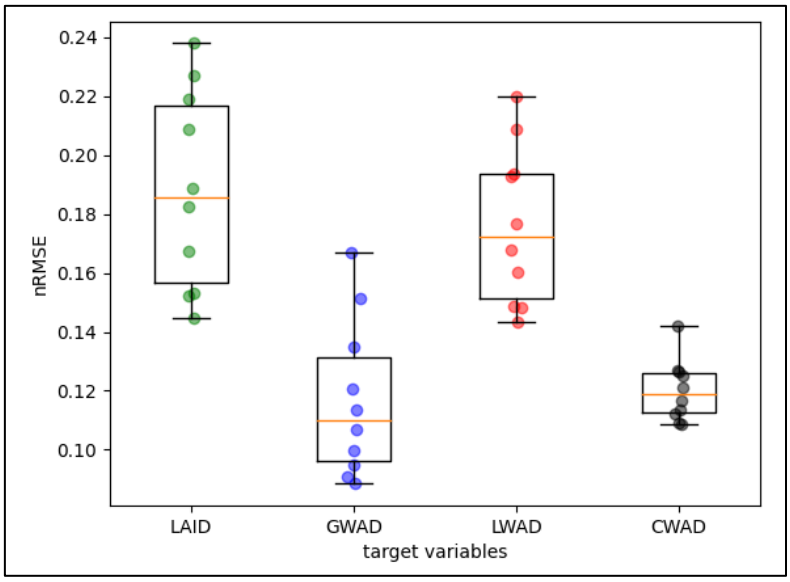
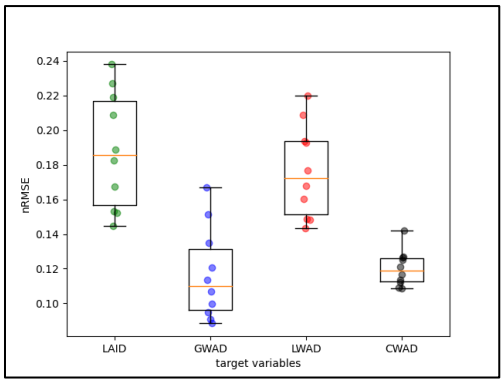
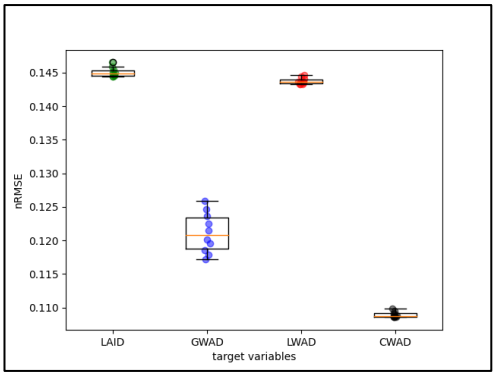


Figure 31:

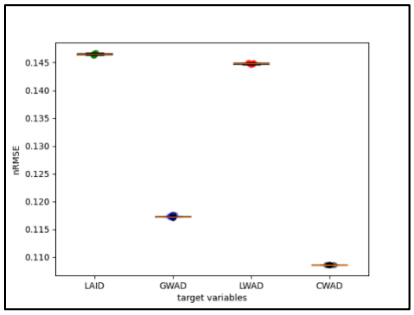
2. *Range reduction example*

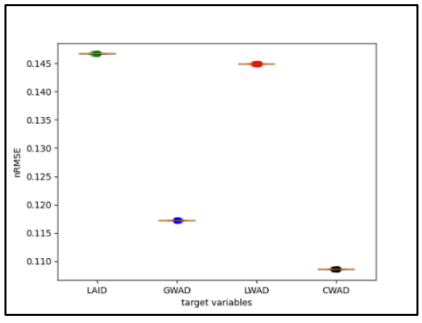
C:\DSSAT48\TSE_workspace -> UFGA7801.SBX_1_1_boxPlot.png and UFGA7801.SBX_1_2_boxPlot.png





C:\DSSAT48\TSE_workspace -> UFGA7801.SBX_1_3_boxPlot.png and UFGA7801.SBX_1_4_boxPlot.png





APPENDIX

A.1 The nRMSE error minimisation method (nRMSE) (Memic et al. 2021)

For quantifying the variation between simulated (Si) and observed (Oi) values the statistical method of nRMSE (Eq. 2) is used. The nRMSE, is RMSE (Eq. 1) normalized by mean ($\overline{O}$) for each observed crop variable.

$$RMSE = \left[\frac{1}{n}\sum_{i=1}^{n}(S_i - O_i)^2\right]^{0.5} \qquad (1)$$

$$nRMSE = \frac{RMSE}{\bar{O}} \qquad (2)$$

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. 3).

$$AVG_{nRMSE(n)} = (GWAD_{nRMSE(n)} + LAID_{nRMSE(n)} + CWAD_{nRMSE(n)} + LWAD_{nRMSE(n)})/4 \quad (3)$$

The selection of the coefficient combination with the lowest nRMSEs average across all target variables proves to be a good solution (Table 1, AVG-nRMSE=0.12).

Table *1* 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.2021).

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

A.2 The nRMSE multiple treatment-based goodness of fit criteria (nRMSE) (Memic et al. 2021)

Cultivar coefficients can be optimized 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 optimization 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 seasons and locations. In Table 2 are shown single treatment "optimums" and multiple treatment "optimum".

For demonstrating single treatment and multiple treatment-based cultivar coefficient optimization 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 localizing "optimums". First single treatment "optimums" are localized based on the lowest average nRMSE (AVG-nRMSE, Table 2, section a, grey fields).

In second step multiple treatment cultivar coefficient "optimum" is localized 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 2, section b, grey fields) with lowest treatment based average nRMSE (Table 2, section b, blue field).

Table 2 '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. (Memic et al. 2021)

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)

A.3 The Relative Weight –nRMSE (RW-nRMSE)

A *Relative Weight -nRMSE* (*RW-nRMSE*) complementary method was developed with relative weights assigned to the selected target variables (1 to 5). A relative weight can be assigned to each target variable separately, where 5 is the most accurate and 1 the least accurate, in relative terms. Each accuracy weight has weight coefficient assigned to it in the TSE code. With relative weights each target variable's accuracy can be selected, giving users more flexibility in terms of target variable prioritization in the cultivar coefficients estimation process. The relative weights multiplier (rw) in Eq. 4 is used to artificially inflate nRMSE of specific target variables. By doing so the size of specific target variable is artificially increased in the overall AVG-

nRMSE that results at the end in cultivar coefficient combination prioritizing those variables over others.

$$AVG_{nRMSE(n)} = (rw_1 * GWAD_{nRMSE(n)} + rw_2 * LAID_{nRMSE(n)} + rw_3 * CWAD_{nRMSE(n)})/3 \quad (4)$$

It will be additionally explained!

A.4 The Rank Weighting –nRMSE (Rank-nRMSE)

A *Relative Weight -nRMSE* (*RW-nRMSE*) complementary method was developed with priority ranks (1 to 5). A rank can be assigned to each target variable separately, where 1 is the most accurate and 5 the least accurate, in relative terms. Each accuracy rank has weight coefficient assigned to it in the TSE code (Table 4), 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 prioritization in the cultivar coefficients estimation process (Table 4), as shown in the following example with GWAD having rank 1 and the other target variables rank 4.

Table 4 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 A.1a). The initial selection criteria were set to 0.0 (Figure A.1a). With the initial selection criteria, the *while* loop (flow controlling statement executed repeatedly until specified conditions are meet in the programming language) is initialized (Figure A.1b). Within the *while* loop, the *if* condition is set for finding an overall statistically acceptable solution (Figure A.1c). 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 4) 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.

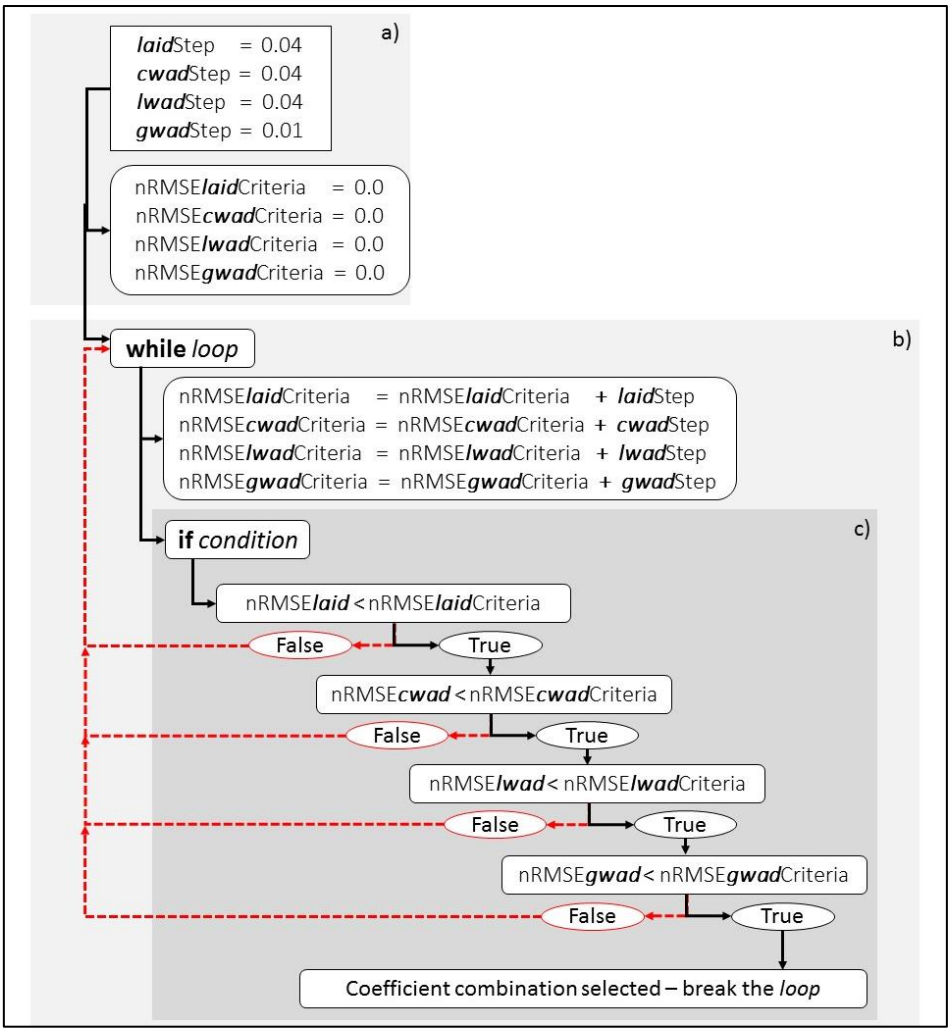


Figure A.1 The RW-nRMSE cultivar coefficient combination selection method (Röll et al. 2020)

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 1.

Table *5* 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 (Memic et al.2021).

	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*

A.5 Range reduction – generating coefficient combinations (Memic 2022)

(This method must be used carefully in case if "Fixed initial Min – Max" check box is NOT checked! It is designed to provide the best statistical fit between simulated and observed. The best statistical fit providing coefficients might not be Physiologically meaningful!)

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 gridding - coefficient variation* (Table 6) can be used to systematically investigate coefficient ranges in search of coefficient values that provide the best statistical fit. For example, P5 coefficient value for minimum 100, maximum 900 and increment step 3.3 can be passed into the cultivar file and for each coefficient value model executed. In this example as shown in Table 6 for a coefficient range from 100 to 900 with increment steps of 3.3 a total 243 coefficient variations are executed for the P5 coefficient with Exhaustive gridding method. To overcome time losses in the process of cultivar coefficients estimation based on the statistical fit (lowest nRMSE), a range reduction method (Röll et al. 2020) was implemented (Table 6). With range reduction method four global phases are conducted in the process of estimating cultivar coefficient with smallest average nRMSE. Greater increment steps are used in the first phase for each given coefficient range with P5 coefficient having minimum 100, maximum 900 and increment step 200. Based on the lowest nRMSE, the value for coefficient is selected, i.e., 300. In the second Phase, the new coefficient ranges with a narrower increment step are executed with P5 having minimum 180, maximum 420 and increment step 60. Based on the lowest nRMSE, the new coefficient "optimum" is selected, i.e. P5=300.In the third phase, the new range for each coefficient is defined with P5 having minimum 270, maximum 330 and increment step 15. In the final phase P5 minimum is 263.2 and maximum 276.7 with increment step of 3.3 is passed into the cultivar file and based on lowest nRMSE P5=263.2 is selected. Based on the range reduction approach, 48% fewer combinations are executed when compared to exhaustive gridding coefficient variations. The range reduction method as described in Table 6 is expected to retain a systematic optimum localization approach (achievable with exhaustive gridding variation) and is expected to provide more realistic values for the coefficients when compared to the random generation of cultivar coefficients for allowed ranges. The range reduction is flexible programmed and will work with different scale coefficients. The method will be described in more detail in an additional publication with corresponding programming code written in python.

Example with CERES-Wheat P5 coefficient (Memic 2022)

Exhaustive gridding		**Range reduction method**							
P5		Phase 1		Phase 2		Phase 3		Phase 4	
Min	100	100		180		270		263.2	
Max	900	900		420		330		267.7	
Inc. step	**+3.3**	**+200**		**+60**		**+15**		**+3.3**	
								263.2	→263.2
								266.5	
				180		270	→270	269.9	
	100	100		240		285		273.3	
	103.3	300	→300	300	→300	300		276.7	
	…	500		360		315			
	→263.2	700		420		330			
	…	900							
	900								
No. comb.	243	5		5		5		5	
Total	**243**	**20**							

References

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), S. 1391–1402. DOI: 10.13031/trans.14432

Memic, E (2022): Combining crop growth models with the Precision Agriculture concept of yield gap analysis to evaluate yield limiting and reducing factors. Uni Kassel, Germany. *Dissertation.*

Röll, G., Memic, E., Graeff-Hönninger, S. (2020): Implementation of an automatic time-series calibration method for the DSSAT wheat models to enhance multi-model approaches. Agronomy Journal 112(5), 3891-3912. https://doi.org/10.1002/agj2.20328

Interface:
The TSE_v2.2 user interface was created in Qt Designer 5
(https://doc.qt.io/qtcreator/index.html)

Programming language:
The TSE_v2.2 algorithm was written in python 3.7
Python Software Foundation. Python Language Reference, version 3.7. Available at http://www.python.org

Windows runnable:
TSE_v2.2 was compiled into windows runnable with Pyinstaller
(https://www.pyinstaller.org/)

TSE_v2.2 algorithm and interface development/setup by Emir Memic.