AgMIP Calibration Protocol: results, implementation and new directions

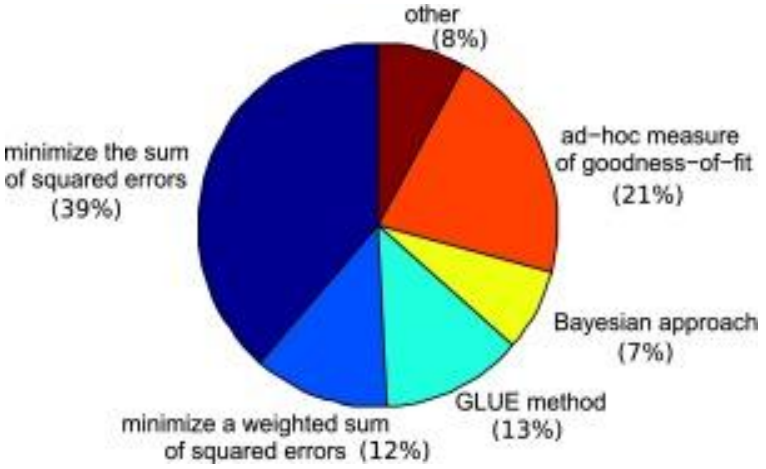
S. Buis (INRAE), *P. Barbillon* (AgroParisTech), *H. Mielenz* (Julius Kühn Institute), *N. Mouhrim* (INRAE), *T. Palosuo* (LUKE), *S. J. Seidel* (BOKU), *P. Thorburn* (CSIRO), *D. Wallach* (INRAE)

XIVth STICS seminar, 17-19 March 2026, organized by ULiege-GxABT & INRAE

Challenges in Crop Model Calibration

Crop model calibration **is still an issue**

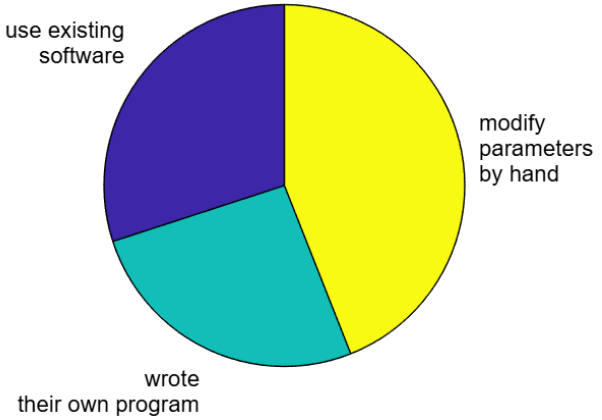
Various practices in the community
(Seidel et al 2018, EJA; **Wallach et al 2021, EMS**)
« We are far from having a consensus on how to calibrate crop models, even in […] relatively simple case »
(Wallach et al 2021, EMS)



Crop model calibration largely impacts the results of simulations

See e.g. Guillaume et al 2011, EJA; Confalonieri et al 2016, EMS;
He et al 2017, AFM; Wallach et al 2021, EJA; Wallach et al 2021, AFM ;
Wallach et al., 2022a, 2022b ; **Wallach et al., 2025, AFM**

Using existing software to calibrate crop models **remains a minority**



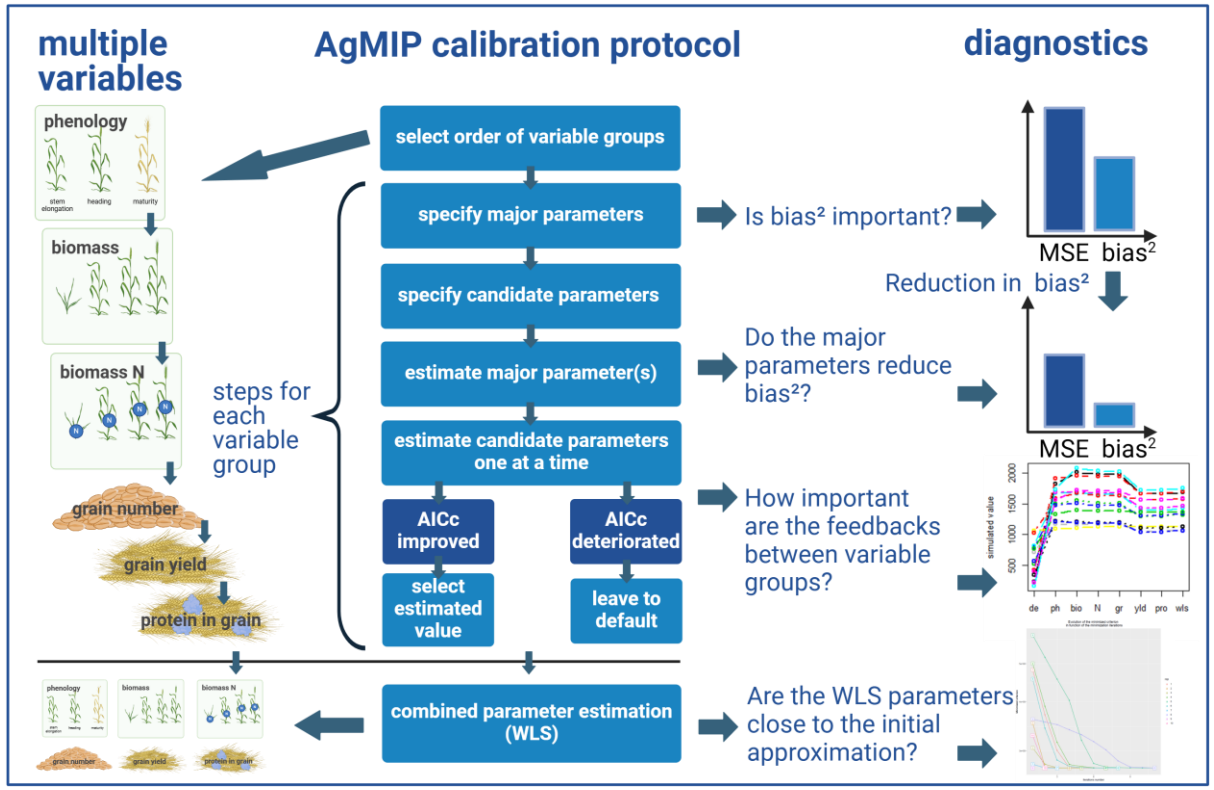
=> Adopting shared, tailored methods and software tools could reduce uncertainties in the calibration process and improve crop models results

The AgMIP Calibration protocol

Wallach et al. 2024, Environ. Modell. Software

Wallach et al. 2025, Eur. J. Agron.

- **Generic**, **comprehensive** calibration with **detailed process documentation**
- Handles **2 common issues** for calibration of complex crop models:
  - **Choice of parameters to estimate**
  - **Use of multiple variables**
- **Multi-step** protocol integrating **expert knowledge** and **automated procedures**



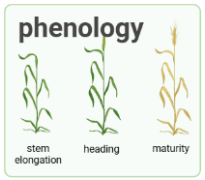
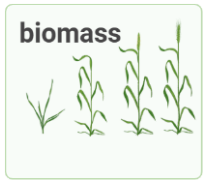
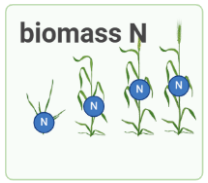
The AgMIP Calibration protocol: steps 1 and 2

- Base default values on knowledge of crop
  - Literature
  - Previous calibration of a similar crop / cultivar

choose default parameter values

- Use all possible observations to achieve a consistent model representation
- Group variables related to the same process and expressed in the same units

list obs. and corresponding sim. variables

Obs. variable	Dates of phenological stages (BBCH30, BBCH55, BBCH90)	Biomass at different dates	N in biomass at harvest	Grain number	Grain yield	Protein content in grain
Sim. variable	iamfs, ilaxs, imats, ...	masec(n)	$\frac{\text{QNplante}}{10 * \text{masec(n)}}$	chargefruit	mafruit	CNgrain * 5.7
Group						

The AgMIP Calibration protocol: step 3

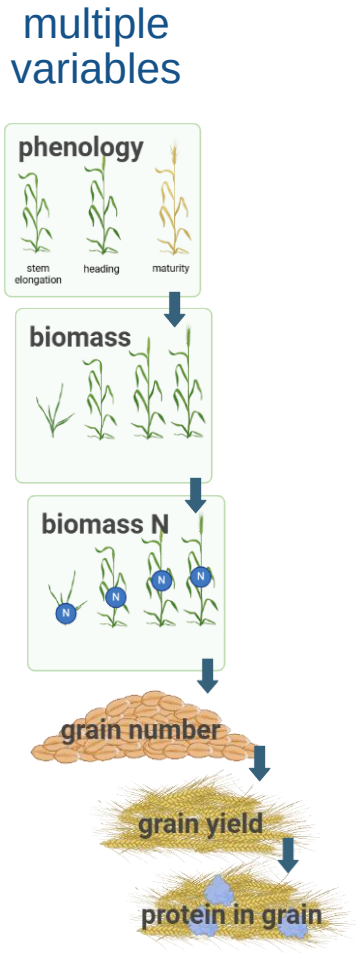
choose default parameter values

list obs. and corresponding sim. variables

select order of variable groups

Order

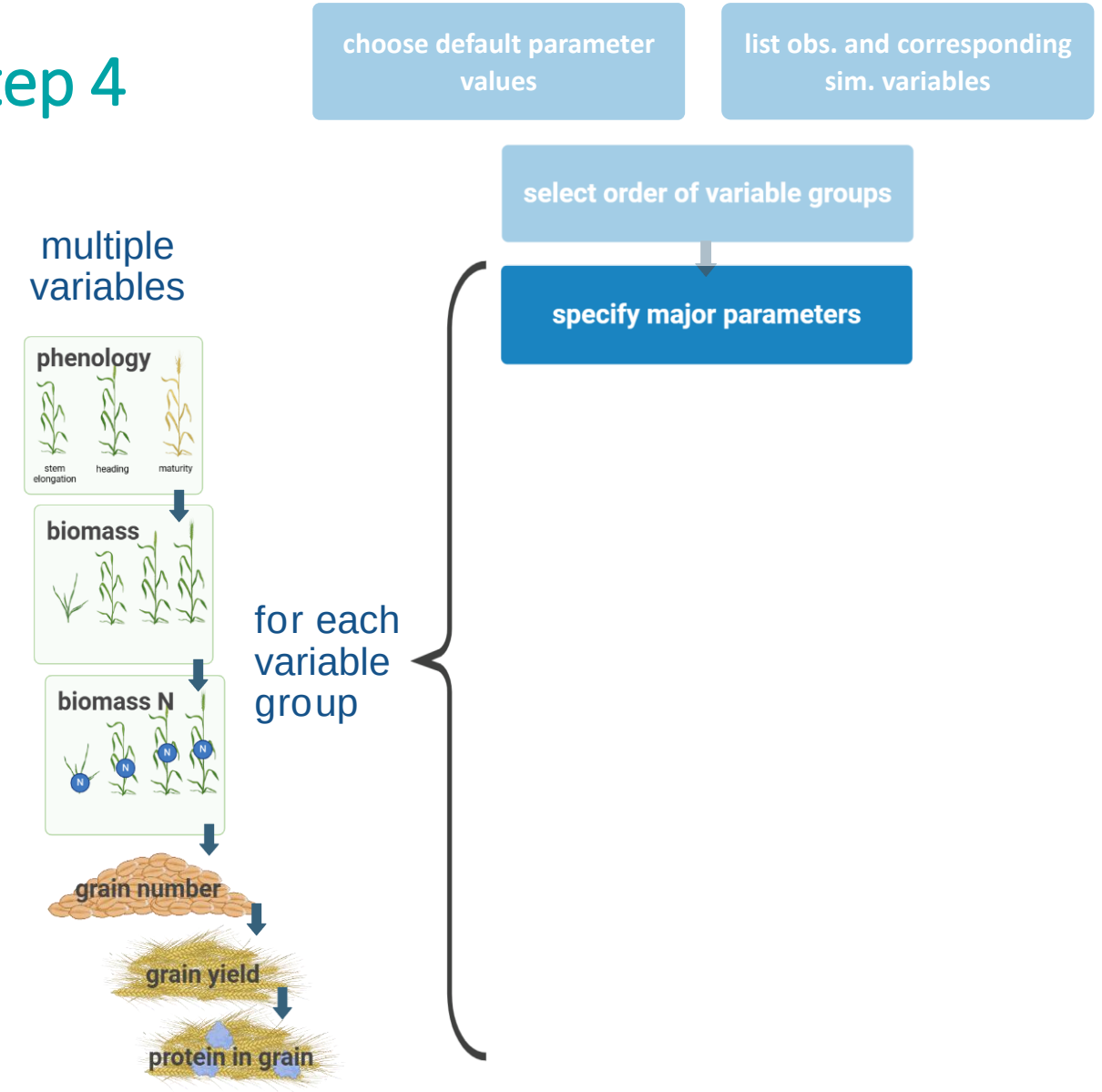
- So that estimating parameters to a later group does not affect previous groups
- Usually, start with phenology



The AgMIP Calibration protocol: step 4

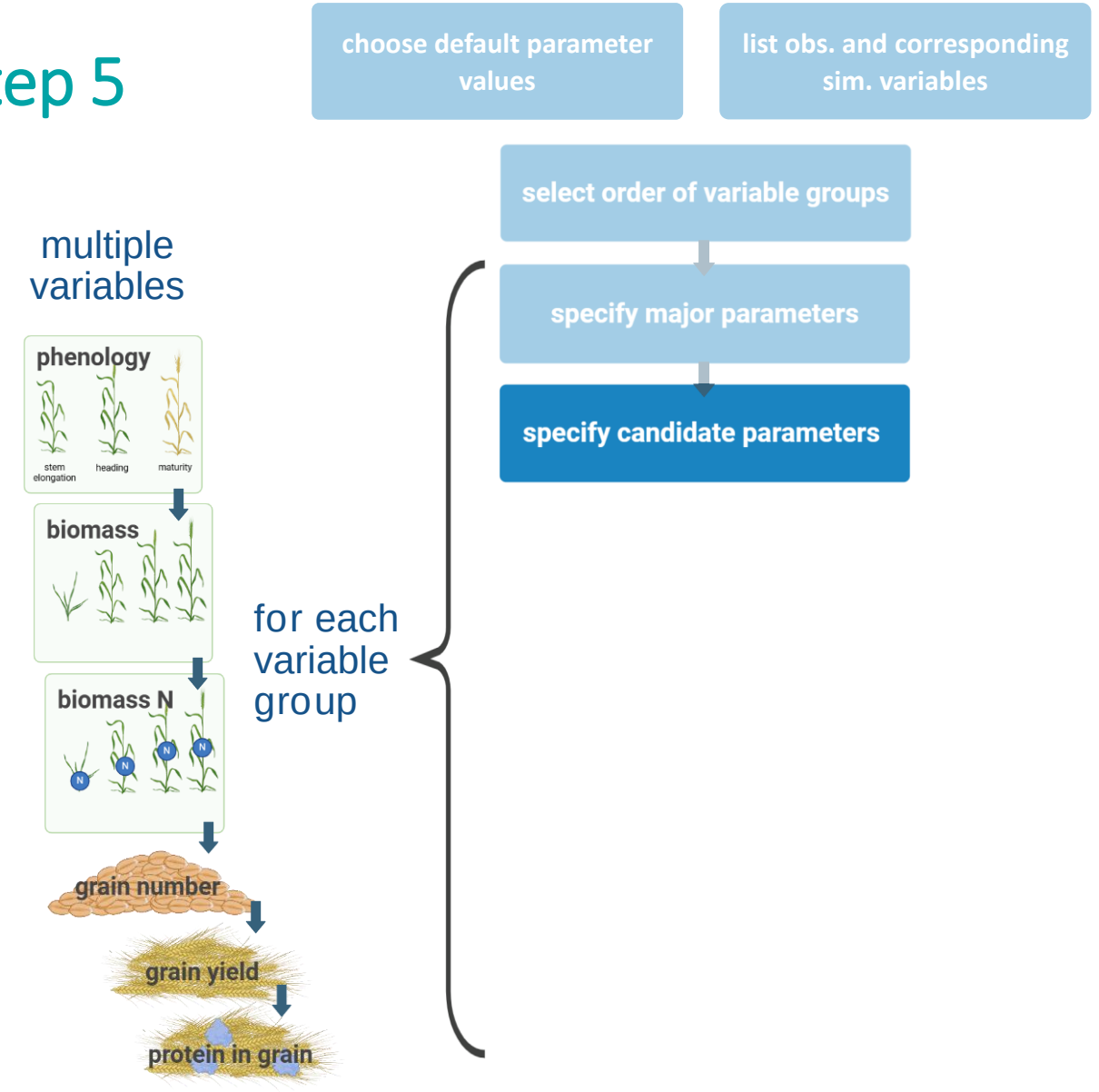
- **Major parameters**: should impact every environment
  - E.g. thermal time to heading.
    Impacts days to heading in every environment
  - E.g. RUE.
    Impacts biomass in every environment
  - This will (nearly) eliminate bias in simulations

- Identify at most:
  - for end-of-season variables: 1 parameter
  - for variables with several in-season data:
    2-3 parameters that affect different periods of the crop cycle, or the level and the temporal dynamics of the variable



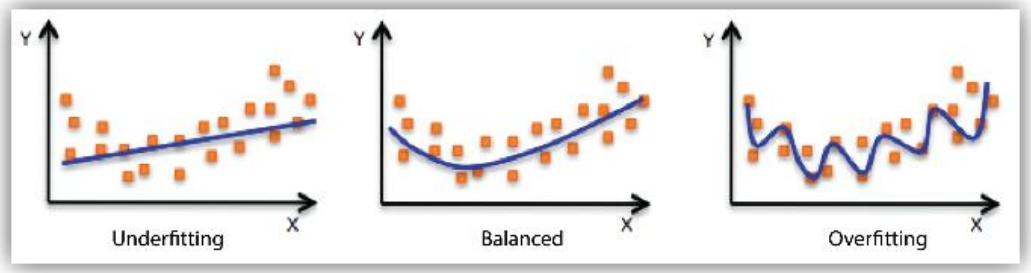
The AgMIP Calibration protocol: step 5

- **Candidate parameters**: likely to explain differences between environments
  - ➢ Don’t choose too many (calculation time)
  - ➢ In order of importance for reducing squared errors

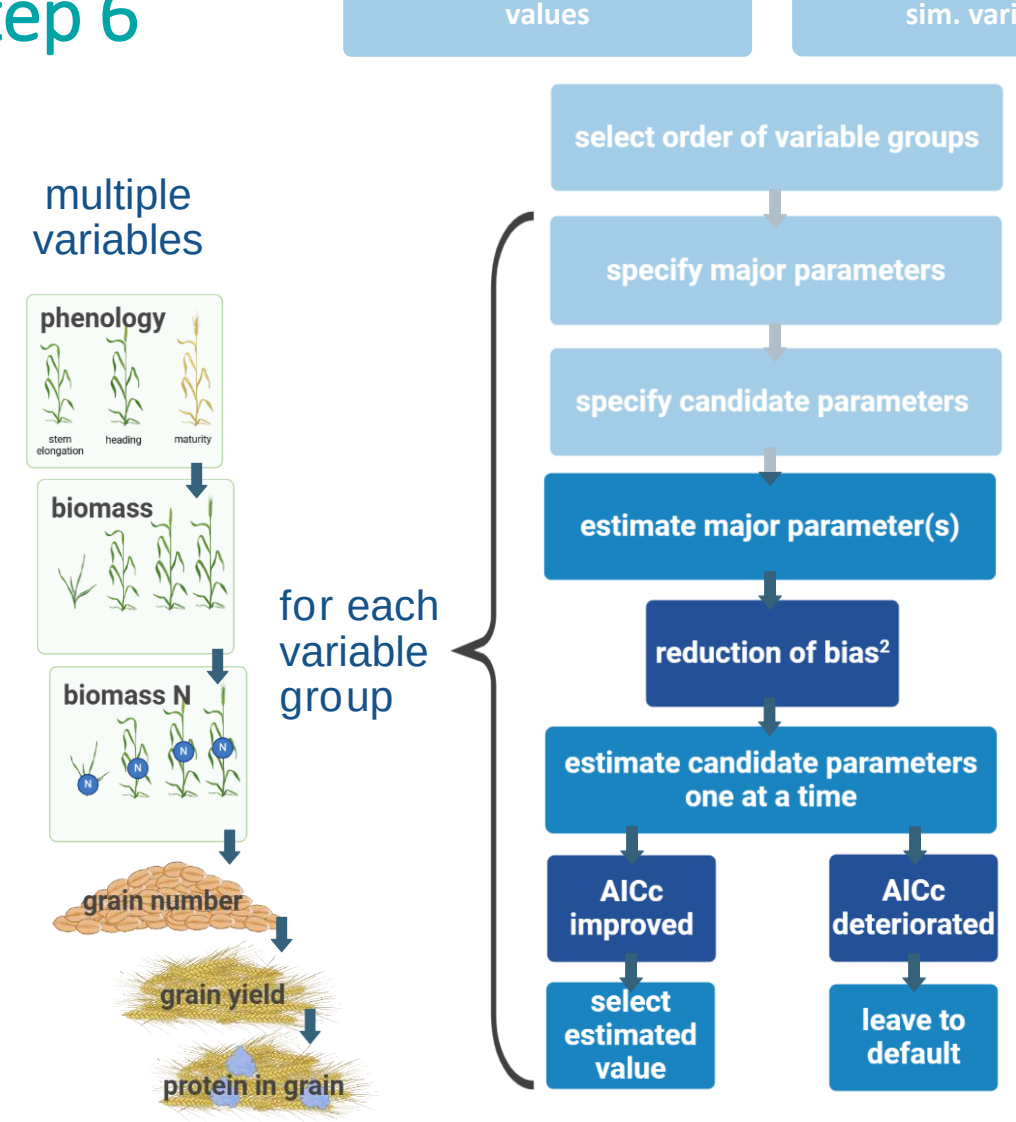


The AgMIP Calibration protocol: step 6

- **Reduce sum of squared errors (OLS) successively for each variable group**
  - e.g. using simplex algorithm with multiple starting points
  - First, estimate bias-reducing parameter(s)
  - Then, consider candidate parameters in turn
    - Use AICc as criterion for **selecting parameters to avoid overfitting**
    - If not selected, parameter retains default value



https://fr.mathworks.com/discovery/overfitting.html

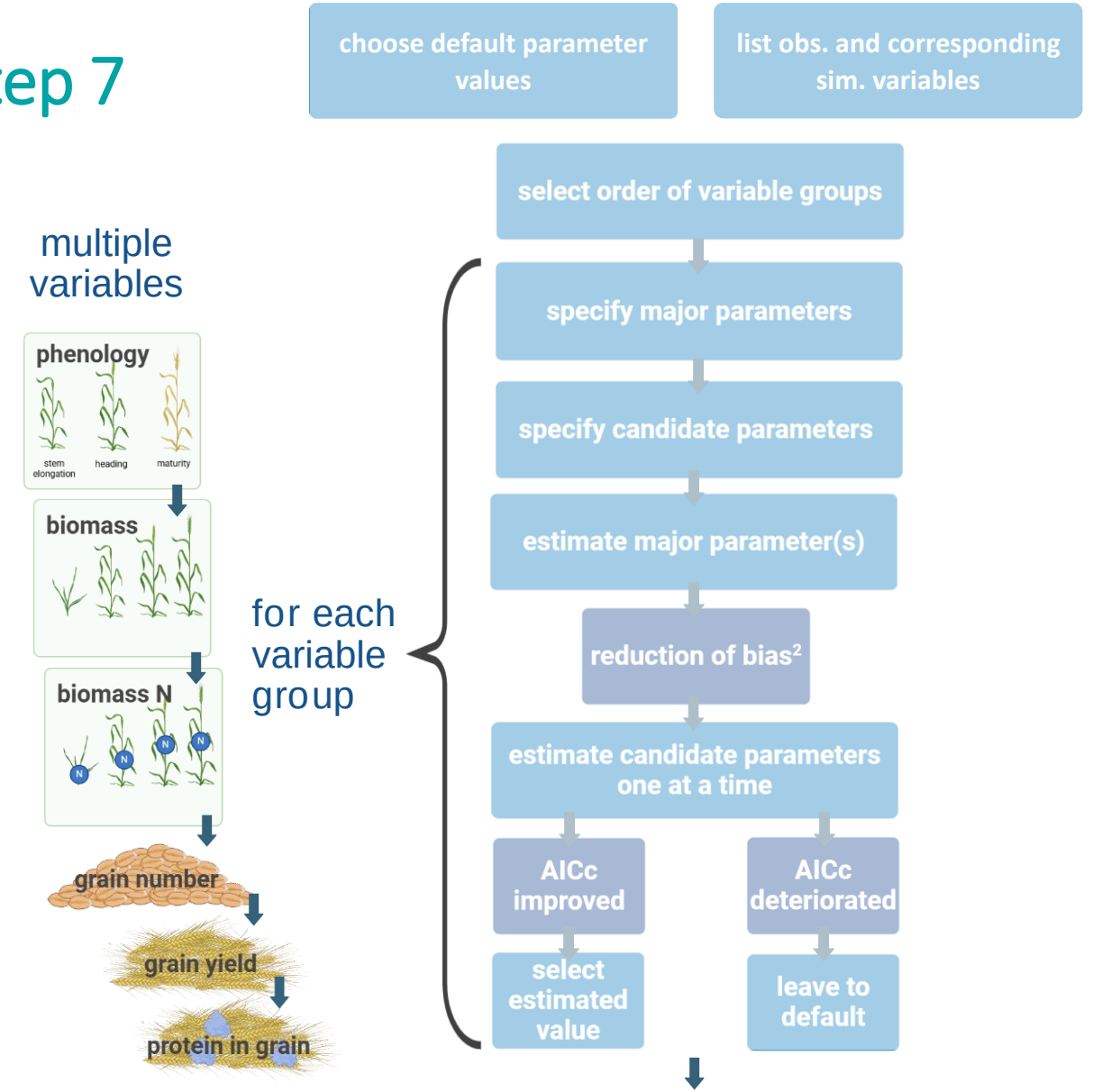


The AgMIP Calibration protocol: step 7

- **Re-estimate all selected parameters together using all observed variables**
  - e.g. using simplex algorithm with many starting points (~20)
  - Use estimates from step 6 as one initial starting point
- Weighted least squares
  - Based on step 6 errors, calculate standard deviation for model error for each group
  - Weight each variable by 1/SD

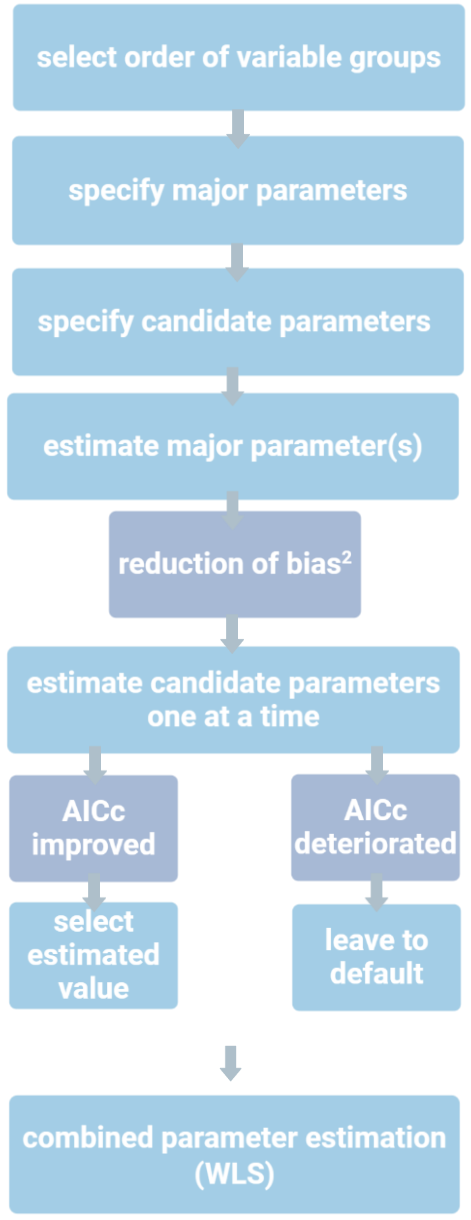
$$WLS = \sum_i \frac{\sum_j \left(y_{ij} - \widehat{y_{ij}}\right)^2}{s_i}$$

where s_i is the standard deviation of model error for group i, estimated following step 6



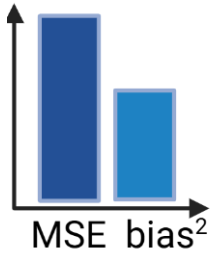
The AgMIP Calibration protocol: step 8

- Use several diagnostic tests
  - Reduction of bias^2
  - Amount of feedback betweeen groups
  - Effect of WLS step
  - Test of overfitting

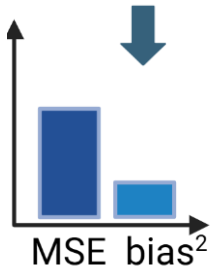


diagnostics

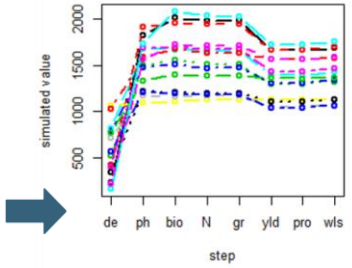
Is bias^2 important?



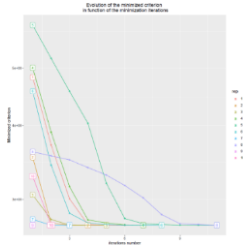
Do the major parameters reduce bias^2?



How important are the feedbacks between variable groups?



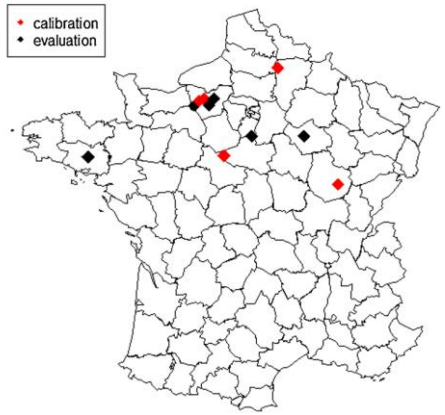
Are the WLS parameters close to the initial approximation?



Evaluation of the protocol in Phase IV multi-model exercise

Data

- Winter wheat variety in France, conventional management, current climate
- 11 study sites over 7 years
- 14 year-location combinations for calibration
- 8 year-location combinations for evaluation
- 9 variables:
  - Dates of BBCH30, BBCH55, BBCH90,
  - Biomass at different dates
  - Number of ears
  - N in biomass at harvest
  - Grain number
  - Grain yield
  - Protein content in grain



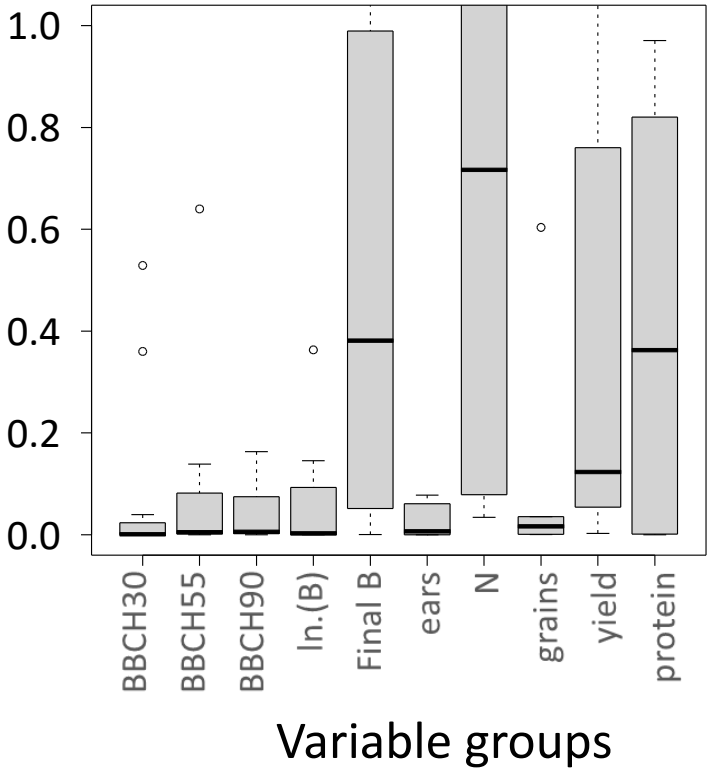
Modeling

- 12 modeling teams
- 8 different modelling structures:
  - DSSAT CERES
  - DSSAT CropSim
  - CropSyst
  - Hermes
  - Hermes2Go
  - DSSAT Nwheat
  - Sirius Quality
  - STICS

Results: bias reduction

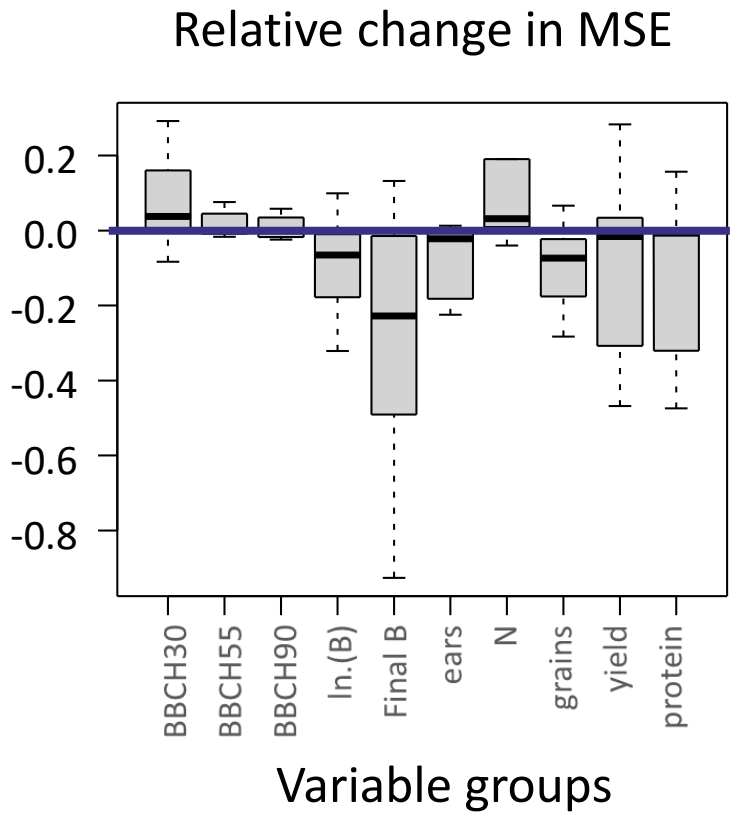
- For most teams and variables bias^2 was substantially reduced
- For yield and biomassN difficult to identify effective bias-reducing parameters
  - ➢ both variables are a result of many interlinked processes
- If bias-reducing parameter for some variable group is ineffective
  - ➢ reconsider the choice
- Perhaps there is none
  - ➢ only candidate parameters

Bias^2 after / bias^2 before step 6 for the respective variable



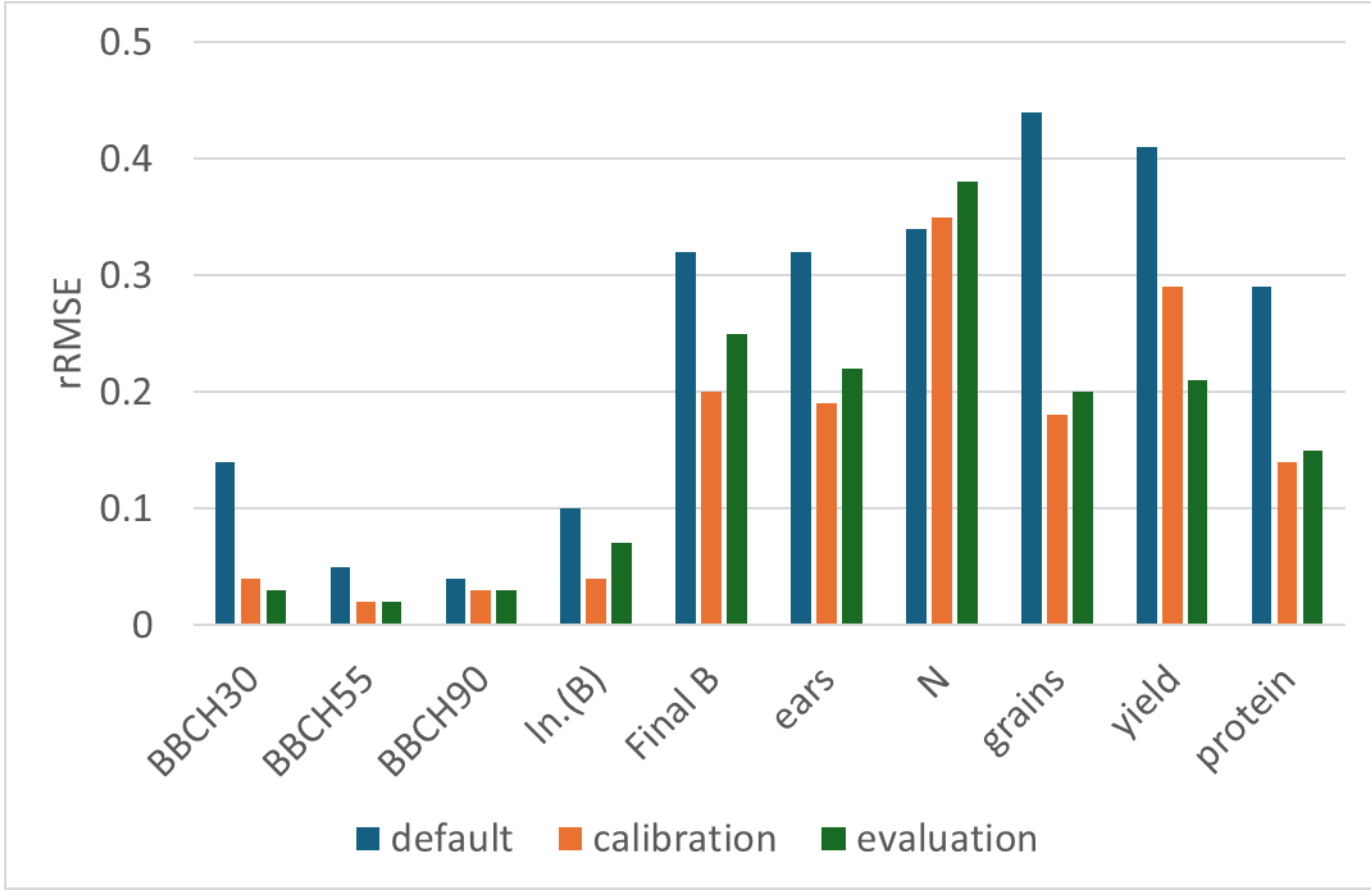
Results: effect of weighted least squares step

- Aim of step 7: redress problems caused by feedbacks
- Give appropriate weightings to the different variables
- In most cases, step 6 gave a good approximation to the parameters for WLS estimation
- **For most variables the WLS step reduced MSE**



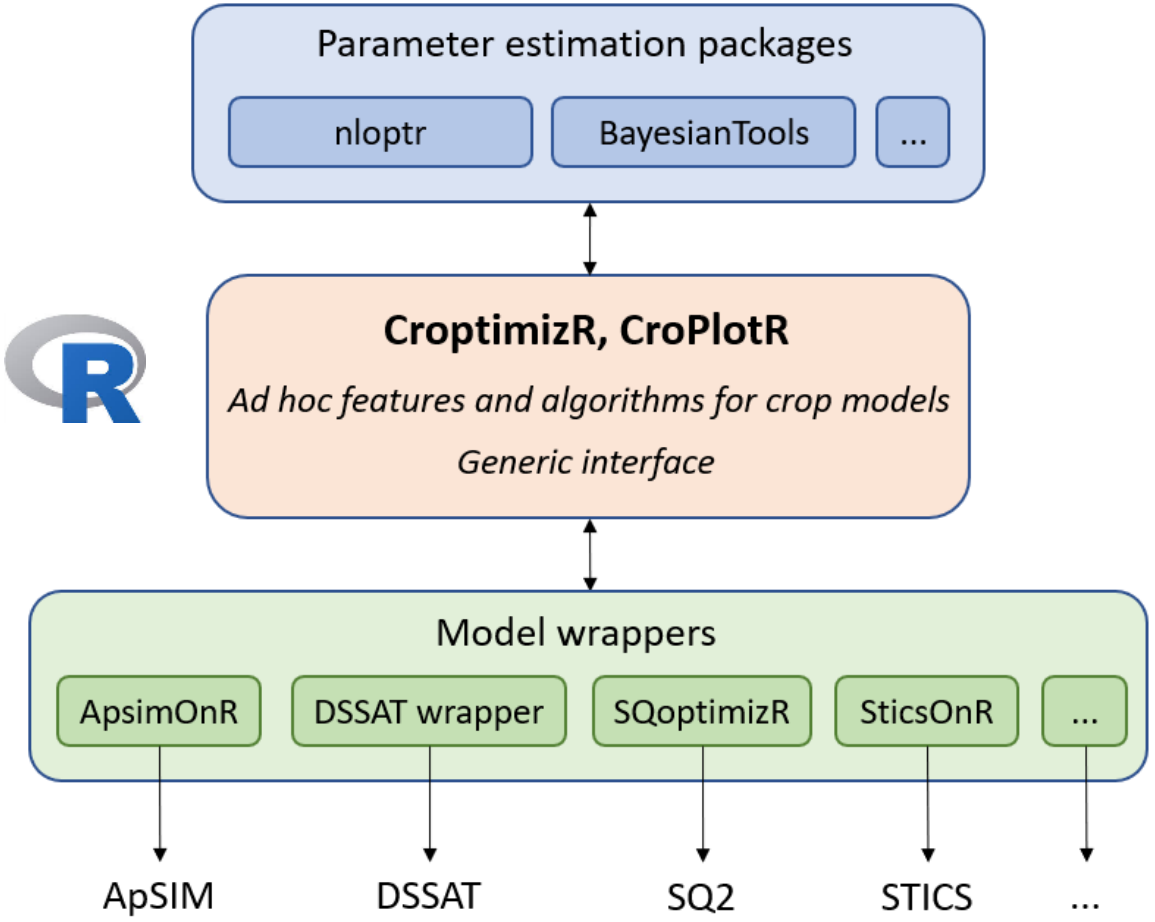
Results: test of overfitting

- rRMSE-values for nearly all variables were reduced by calibration
- rRMSE-values of evaluation data were on average slightly higher than those of calibration data
  - ⇒ Little or no overfitting
  - ⇒ Use of AICc-criterion was efficient



Implementation: the CroptimizR and CroPlotR packages

A Unified Framework for Crop Model Calibration and Evaluation



Implement **automated calibration** and **evaluation methods** for crop models within a **unified framework**

⇒ **Facilitate the comparison of methods across different crop models**

⇒ Combine efforts to **deliver valuable and easily applicable methods** for all crop model users

Implementation: the AgMIP calibration protocol

Protocol description xls file

Group definition and order

variable	group
Date_BBCH30	phenology
Date_BBCH55	phenology
Date_BBCH90	phenology
Biomass	plant_biomass
Grain_Number	grain_number
Grain_Yield	yield

Major parameters

parameter	group	default_value	lower_bound	upper_bound
stlevamf	phenology	324.8	150	400
stamflax	phenology	446.8	150	500
stdrpmat	phenology	820	500	900
efcroiveg	plant_biomass	5.3	3	6
efcroirepro	plant_biomass	3.5	3	6
cgrain	grain_number	0.0324	0.03	0.04
vitircarbT	yield	0.00031	0.00005	0.002

Candidate parameters

parameter	group	default_value	lower_bound	upper_bound
jvc	phenology	58.364	25	60
sensrsec	phenology	0.8	0	1
belong	phenology	0.0228	0.005	0.03
jvcmini	phenology	11.8	2	15
dlaimaxbrut	plant_biomass	0.003188	0.000005	0.005
durvieF	plant_biomass	260	40	300
vlaimax	plant_biomass	2.38	1.5	2.5

load_protocol_agmip(protocol_file_path)

```
$lb
   stlevamf    stamflax    stdrpmat   efcroiveg efcroirepro
    1.5e+02     1.5e+02     5.0e+02     3.0e+00     3.0e+00
$ub
   stlevamf    stamflax    stdrpmat   efcroiveg efcroirepro
    4.0e+02     5.0e+02     9.0e+02     6.0e+00     6.0e+00
$default
   stlevamf    stamflax    stdrpmat   efcroiveg efcroirepro
 324.800000  446.800000  820.000000     5.300000    3.500000
```

```
$phenology
$phenology$major_param
[1] "stlevamf" "stamflax" "stdrpmat"
$phenology$candidate_param
[1] "jvc"       "sensrsec"  "belong"    "jvcmini"   "stressdev"
$phenology$obs_var
Date_BBCH10 Date_BBCH30 Date_BBCH55 Date_BBCH90
    "ilevs"     "iamfs"     "ilaxs"     "imats"

$plant_biomass
```

run_protocol_agmip(obs_list, model_wrapper, model_options, param_info, step, ...)

Implementation: What does run_protocol_agmip return?

A single structured R object summarizing the calibration protocol results

Final estimated values of parameters

stlevamf	stamflax	stdrpmat	stressdev	efcroiveg	efcroirepro	dlaimaxbrut	vlaimax	Vmax2
2.0e+02	3.6e+02	7.0e+02	6.3e-03	4.1e+00	3.4e+00	4.2e-04	2.3e+00	5.2e-02

Values of parameters after each step

name	default	step6	step7
stlevamf	3.25e+02	2.02e+02	2.03e+02
stamflax	4.47e+02	3.59e+02	3.59e+02
stdrpmat	8.20e+02	6.96e+02	7.04e+02
stressdev	6.00e-01	3.72e-03	6.31e-03
efcroiveg	5.30e+00	3.96e+00	4.09e+00

Statistics on variables after each step

step	variable	Bias2	MSE	rRMSE	EF
Default	masec_n	4.38e-02	1.01e-01	0.19189	0.919
Step6.phenology	masec_n	3.50e-01	4.09e-01	0.38580	0.671
Step6.plant_biomass	masec_n	1.29e-05	1.99e-02	0.08507	0.984
Step6.plant_N_content	masec_n	7.54e-03	2.55e-02	0.09644	0.979
Step6.grain_number	masec_n	7.54e-03	2.55e-02	0.09644	0.979
Step6.yield	masec_n	6.08e-03	2.55e-02	0.09643	0.979
Step6.seed_protein	masec_n	6.08e-03	2.55e-02	0.09643	0.979
Step7	masec_n	6.98e-04	1.69e-02	0.07840	0.986

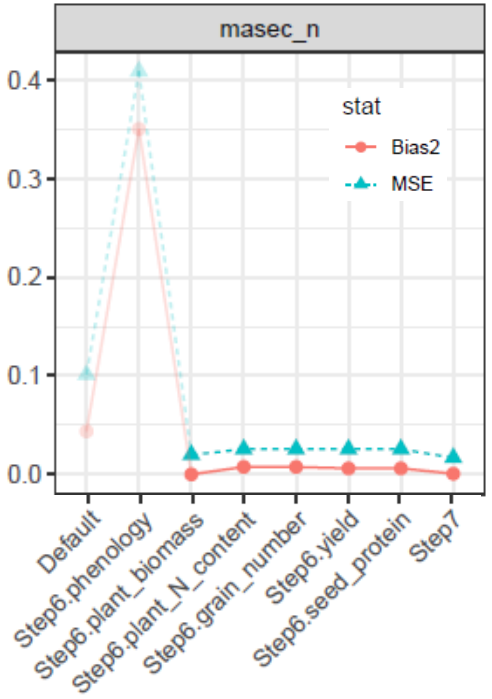
Summary results for Step 6 and 7

Full outputs of all calls to estim_param in Step 6 and 7, saved to disk
(detailed optimization and parameter selection results, Nelder-Mead simplex diagnostics …)

Implementation: What does run_protocol_agmip return?

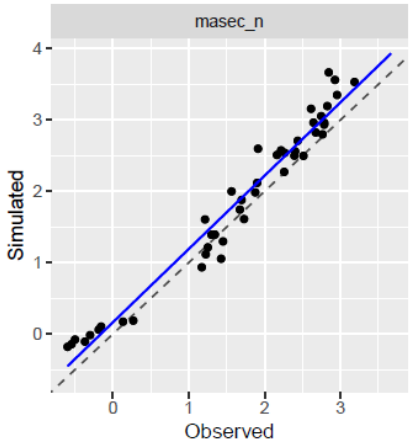
Built-in diagnostics plots for the protocol

MSE and bias² per variable after each step

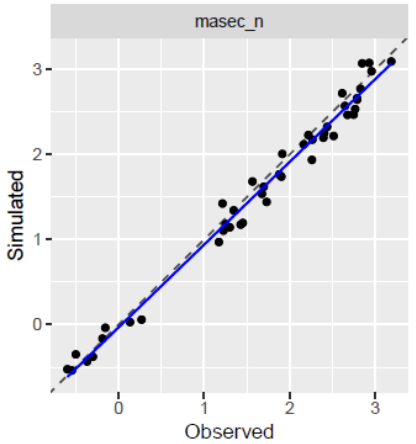


Simulated VS observed values per variable after each step

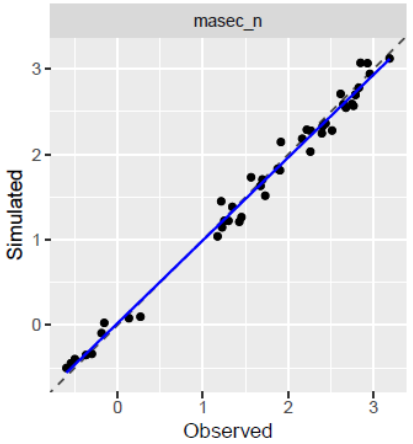
Default



Step6



Step7



Conclusions

- Protocol **based on statistical theory**
- Proposes **practical solutions for parameter selection and use of multiple variables**
- Applicable to **any crop model and dataset**
- **Robust** in practice and worked **consistently** across 12 models and teams
  - For most teams and variables **bias² was substantially reduced**
  - **Avoidance of overfitting** through selection of estimated parameters
- **Implementation available** in CroptimizR that automates the protocol and generates diagnostics
- Associated with **detailed process documentation**

⇒ **A new step toward automated and reproducible crop model calibration**

Contents lists available at ScienceDirect

European Journal of Agronomy

journal homepage: www.elsevier.com/locate/eja

Towards improved calibration of crop models – Where are we now and where should we go?

S.J. Seidel[a,*], T. Palosuo[b], I

[a] Institute of Crop Science and Resource Conse
[b] Natural Resources Institute Finland (Luke),
[c] CSIRO Agriculture and Food, Brisbane, Aust
[d] INRA, UMR AGIR, Castanet Tolosan, Franc

Contents lists available at ScienceDirect

Agricultural and Forest Meteorology

journal homepage: www.elsevier.com/locate/agrformet

How well do crop modeling groups predict wheat phenology, given calibration data from the target population?

Daniel Wallach [a], Taru Palosuo [b], Peter Thorburn [c], Emmanu
Bruno Basso [f], Samuel Buis [g], Neil Crout [h], Camil
Thomas Gaiser [k], Cécile Garcia [d]

Contents lists available at ScienceDirect

Environmental Modelling and Software

journal homepage: www.elsevier.com/locate/envsoft

Multi-model evaluation of phenology prediction for wheat in Australia

Daniel Wallach [a], Taru Palosuo [b,*], Peter Thorburn [c]
Senthold Asseng [e], Bruno Basso [f]
Roberto Ferrise [i]

Contents lists available at ScienceDirect

Environmental Modelling and Software

journal homepage: www.elsevier.com/locate/envsoft

The chaos in calibrating crop models: Lessons learned from a multi-model calibration exercise

Taru Palosuo [b,1], Peter Thorburn [c,1], Zvi Hochman [c], Emmanuelle Gourdain [d],
Asseng [ak], Bruno Basso [f], Samuel Buis [g], Neil Crout [h],
Ferrise [i], Thomas Gaiser [k], Cecile Garcia [d],
[n], Steven Hoek [o], Heidi Horan [c],
Erik Jansson [r], Qi Jing [s],

Agronomy for Sustainable Development (2023) 43:46
https://doi.org/10.1007/s13593-023-00900-0

Thanks for your attention!

A calibration protocol for soil-crop models

Taru Palosuo [d,*], Peter J. Thorburn [e],
Dumont [j],

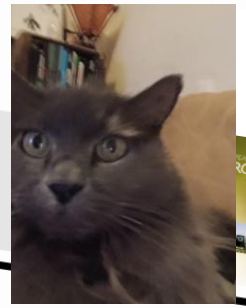
Contents lists available at ScienceDirect

European Journal of Agronomy

journal homepage: www.elsevier.com/locate/eja

uating the AgMIP calibration protocol for crop models; case study and diagnostic tests

Wallach [a,1], Kwang Soo Kim [b], Shinwoo Hyun [b], Samuel Buis [c], Peter Thorburn [d]
Mielenz [e], Sabine Julia Seidel [a,f,*], Phillip D. Alderman [g]
ad Hassan Fallah [i], Gerrit Hoogenboom [j,k]
stian Kersebaum [m,n,o]



Contents lists available at ScienceDirect

Agricultural and Forest Meteor

journal homepage: www.elsevier.com/locate/agrf

Proposal and extensive test of a calibration protocol for crop phenology models

Daniel Wallach [1]

Why is there so much variability in crop multi-model stu

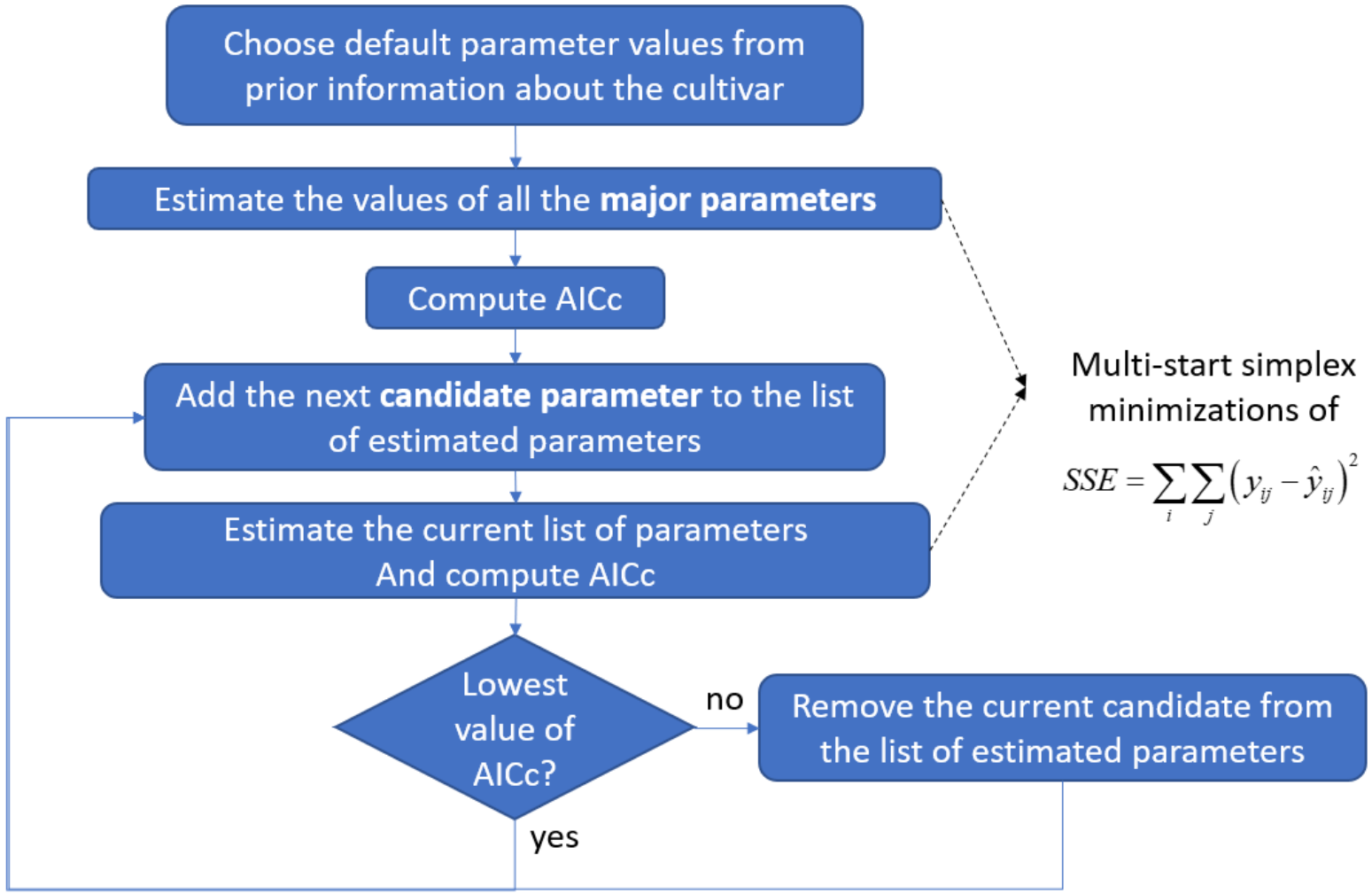
Daniel Wallach [a], Taru Palosuo [b,*], Henrike Mielenz [c], Samuel Buis [d]
Senthold Asseng [f], Benjamin Dumont [g], Roberto Ferrise [h], Sebastian
Afshin Ghahramani [aa], Matthew Tom Harrison [j], Zvi Hochman [e], Ge
Mingxia Huang [l], Qi Jing [m], Eric Justes [n], Kurt Christian Kersebaum [o],
Elisabet Lewan [s], Ke Liu [j], Qunying Luo [t], Fasil Mequanint [i,1], Claas

AgMIP calibration

CroptimizR

Phase III (2021-2022): let's do the same but with a common methodology

Step 5



Example for the STICS model, French dataset

Estimated parameters	Sum of squared errors	BIC
stlevamf, stamflax	405	81.47
stlevamf, stamflax, jvc	349	80.64
stlevamf, stamflax, jvc, sensrsec	322	81.71
stlevamf, stamflax, jvc, belong	349	83.97
stlevamf, stamflax, jvc, jvcmini	319	81.45
stlevamf, stamflax, jvc, stressdev	349	83.97

Total cost : (n° of candidates + 1) = 6 estimations

(Forward regression : 28 estimations,
All combinations : 127)

Results: amount of feedback

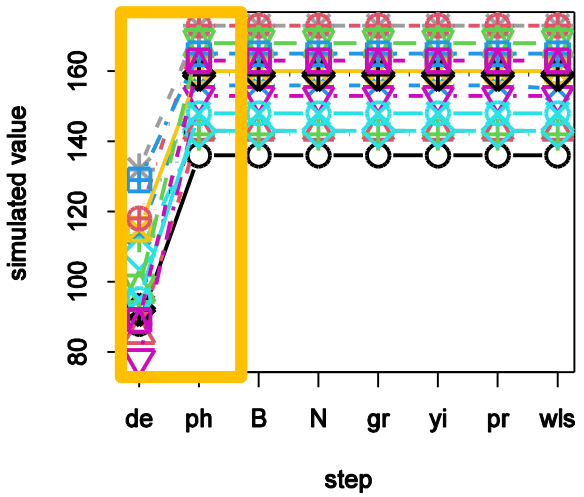
Typical order was: phenology – biomass – ears – biomass N – grain number – yield – grain protein

- No feedback for phenology
- Feedback for other parameters:
  - Biomass and ears

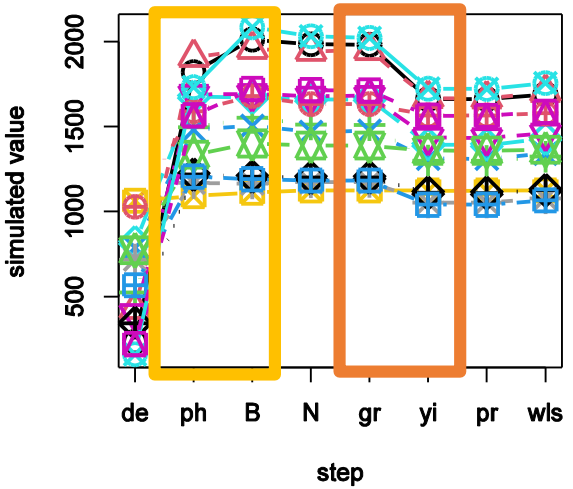
➢ Little feedback leads to a good first approximation of the WLS parameters

Example: NWHEAT DSSAT

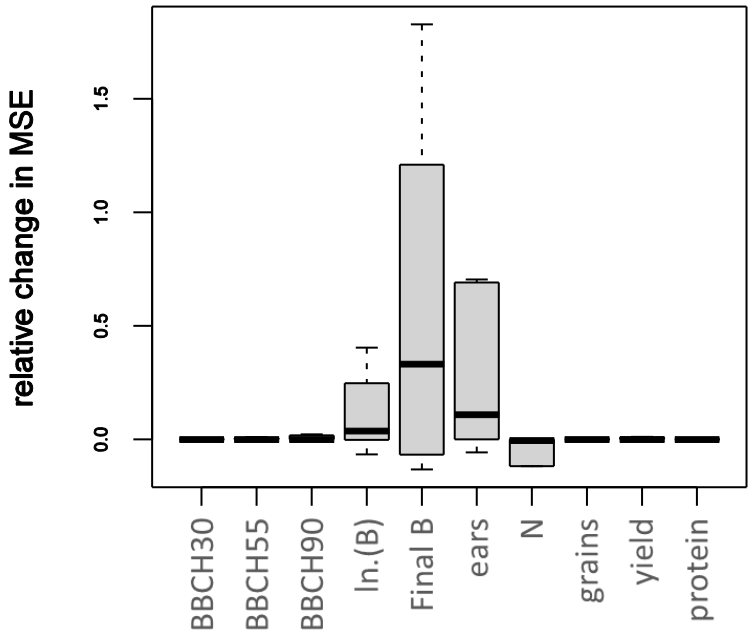
BBCH30



Biomass



Feedback for all models



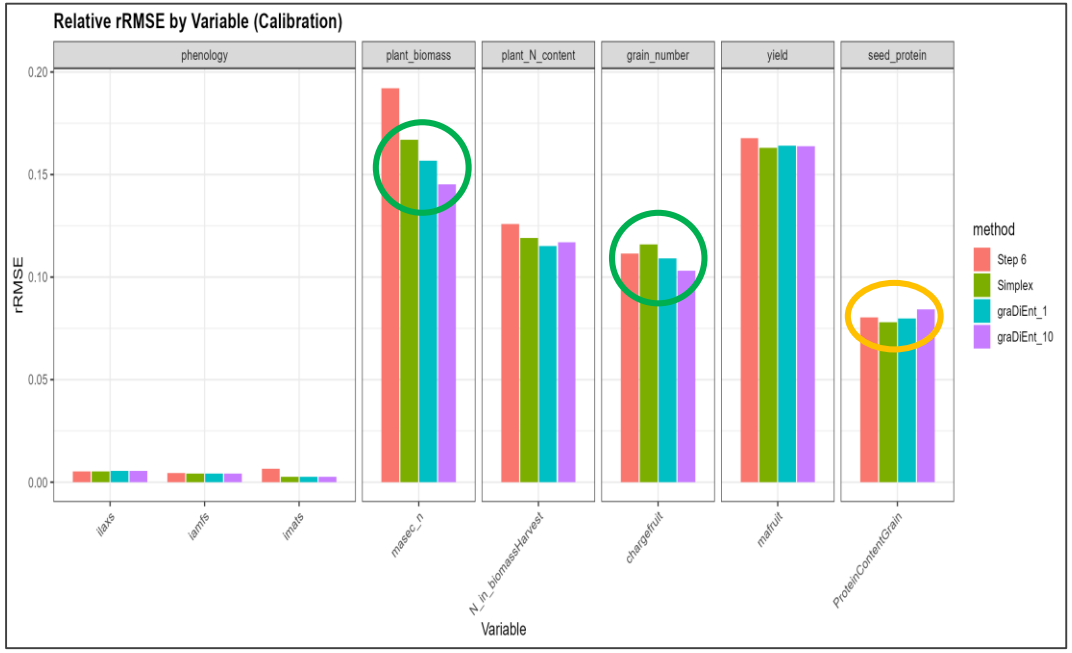
Difficulty of the Step 7 Optimization Problem

- High-dimensional problem (13-30 parameters, 7-10 variables across all models)
- Strong instability of estimated parameters across simplex repetitions
- Large variability of objective values across simplex repetitions

Simplex gets trapped in local minima

Test of Global Optimization Algorithms for step 7

Fit to observed variables on the Calibration dataset

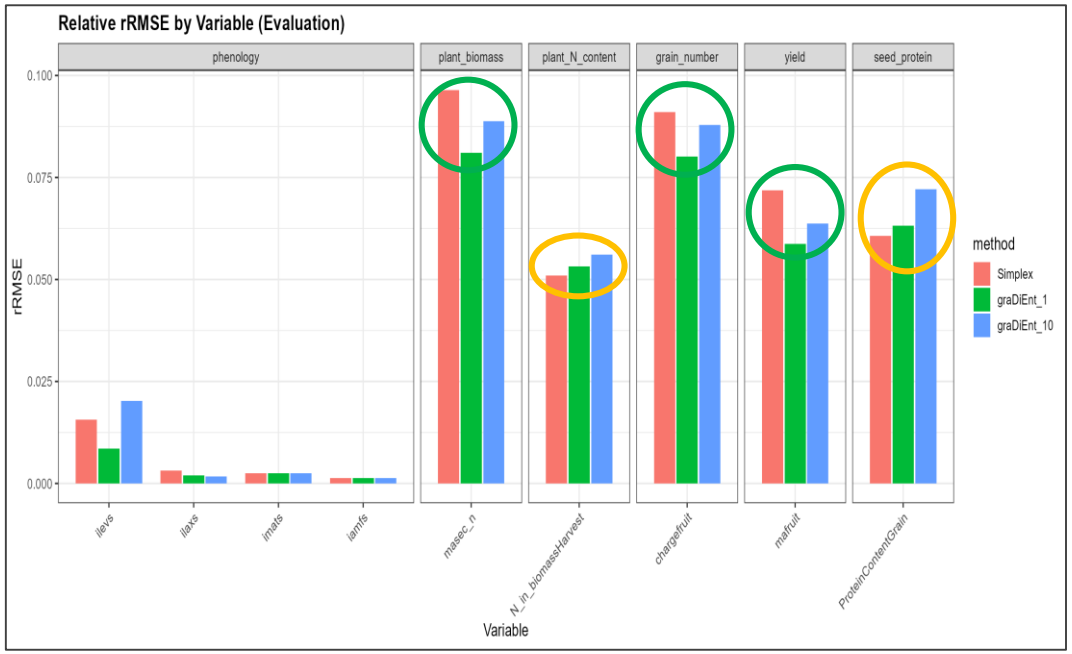


graDiEnt achieved:

- Lower errors for: biomass and grain_number
- Slightly higher errors for: ProteinContent
- Same error for the other variables.

Improvement of maximum 5% of rRMSE in average

Fit to observed variables on the Evaluation dataset



graDiEnt achieved:

- Lower error for: Biomass, Grain_number and Grain_Yield
- Higher error for: proteinContentGrain and plant N content
- Same error for the other variables.

No improvement in average with gradient despite many iterations