

FACCE-MACSUR

Identification and quantification of differences between models

D. Wallach (P206)*; M. Rivington (P150)[†]

Partners involved: Martre (INRA)

* INRA, Toulouse, France

daniel.wallach@toulouse.inra.fr

[†]The James Hutton Institute, Craigiebuckler, Aberdeen. AB15 8QH. UK.
mike.rivington@hutton.ac.uk

Deliverable type: Report

File name:

Deliverable reference num.: D-C.4.2.2

Revision	Changes	Date
1.0	Final version	2014-12-15

Instrument:	Joint Programming Initiative
Topic:	Agriculture, Food Security, and Climate Change
Project:	Modelling European Agriculture with Climate Change for Food Security (FACCE-MACSUR)
Due date of deliverable:	
Submission date:	2014-12-15
Start date of project:	1 June 2012
Duration:	36 months
Deliverable lead partner:	INRA (P206)
Revision:	1.0
Work Package:	CropM 4.2
Document ref number:	

Abstract/Executive summary

A major goal of crop model inter-comparison is model improvement, and an important intermediate step toward that goal is understanding in some detail how models differ, and the consequences of those differences. This report is intended as a first attempt at describing possible techniques for relating differences between model outputs to specific aspects of the models.

Table of contents

Identification and quantification of differences between models	i
Abstract/Executive summary.....	1
Introduction.....	2
Phenology models.....	2
<i>Examine equations.....</i>	<i>2</i>
<i>Create and compare multiple models.....</i>	<i>2</i>
<i>Disentangle differences due to parameters and equations</i>	<i>3</i>
Full crop models.....	4
<i>Examine equations.....</i>	<i>4</i>
<i>Test multiple versions of a specific aspect of a crop model</i>	<i>4</i>
<i>Compare models for contexts chosen to highlight a specific type of response</i>	<i>4</i>
<i>Use multiple outputs to separate different contributions to model variability</i>	<i>4</i>
<i>Identifying families of models.....</i>	<i>5</i>
Conclusions	5
References.....	6

Introduction

A fairly recent tool of crop modelling is multi-model ensemble (MME) simulations, where multiple models are given the same information and are used to run the same simulations. A major objective of these studies is model improvement, through comparison of the different models between themselves and with data. However, it is far from clear how to move from model inter-comparison to model improvement. It is straightforward to ascertain that the simulated values differ among models. Ascribing those differences to specific differences between models is however difficult, and even more difficult is deciding which of the model formulations is best.

In this report we discuss the problem of identifying and quantifying differences between models, and relating those differences to differences in simulated outputs. The differences of interest here may be in specific equations, specific parameter values, general modelling formulations of particular processes, or in fact any level of detail below that of the overall model. The recent studies with MMEs have all posed more or less explicitly this question, of how to relate differences in model outputs to differences in model formulation and/or parameters. However, there are as yet no definitive guidelines on how best to do this, nor even a compendium of possible approaches. This report is intended as a first attempt at describing possible techniques for relating differences between model outputs to specific aspects of the models.

The studies to date may be usefully divided into two categories: studies on models of phenology, and studies on overall crop growth models. There are several reasons that have led to a particular interest in phenology models. In many models phenology is not affected by crop growth, and so can be studied independently of the rest of the crop model. Furthermore, phenology models are usually quite simple, and much easier to work with and analyse than overall crop models. Also, correctly simulating phenology is crucial for correctly simulating growth. Finally, a major impact of climate change will be on phenology, because of its sensitivity to temperature.

In the following, we first consider how to identify and quantify differences between phenology models, and then in the following section consider the problem for full crop models.

Phenology models

Examine equations

The most direct approach to model inter-comparison is to compare directly the equations in the different models. This is usually feasible for phenology models, since they generally have the same form, relating development time to temperature, and possibly photoperiod. This is clearly an important approach, but in general one also wants to go further, and to quantify the consequences on outputs of differences in the equations.

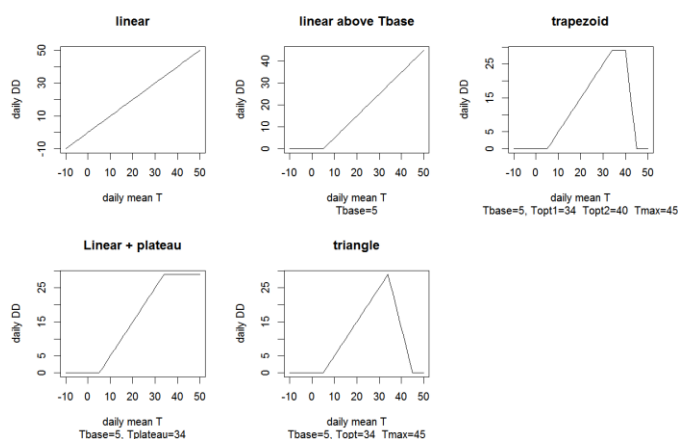
Create and compare multiple models

A straightforward approach is simply to formulate multiple possible models, use them to simulate for a range of contexts, and then compare between models and with observed data. In (Kumudini et al. 2014), 8 different thermal functions were tested against data for more than 1000 maize hybrids in over 50 geographic locations. It was found that different models had different levels of precision, and that precision was in the order calendar days < empirical linear < process based < empirical nonlinear.

This study highlights a major problem in model comparison. Despite the very large data set, few of the locations had supra-optimal temperatures, so that the model comparison does not reliably indicate how the different models will perform under such conditions. The general lesson here is that the conclusions of model inter-comparison or comparison with data may not apply outside the range of contexts studied.

Disentangle differences due to parameters and equations

The above comparison confounds to a certain extent differences in functional form and differences in parameter values, since the phenology models tested had fixed values for the cardinal temperatures (often base temperature, optimum temperature and maximum temperature). In a different, unpublished study, different functional forms were compared where the cardinal temperatures were estimated from data. In this case one is comparing explicitly different functional forms, each with parameter values optimized for the calibration data. The data here are for a single wheat variety over a large range of temperatures, obtained by using multiple planting dates and supplemental heating (Ottman et al. 2012). These data are used in (Asseng et al. 2014). The functional forms considered are shown graphically below.



The results of the fit to the data for the different models are shown in the table below. The linear above Tbase, linear plus plateau and triangle functions all have the same standard error. A difficulty similar to that noted previously arises here. Despite the experimental design, which was aimed specifically at exploring high temperatures, there is not enough high temperature data to discriminate clearly among functional forms that differ in their response at high temperatures. Therefore there is no advantage, in terms of reduced standard deviation, to adding parameters that describe high temperature behaviour. Once again, this is related to the contexts that are simulated. If higher temperature contexts were simulated, there would probably be clearer differences between the functional forms.

An important conclusion here is the importance of considering both the effects of functional form and parameter values when comparing models. In this example, several different functional forms show the same behaviour, once they are provided with optimized parameter values.

model	residual standard error and optimized parameter values
Linear model	9.54
DD anthesis	1098
DDmaturity	1866
Linear above Tbase	3.77
Tbase	5.20
DD anthesis	737
DD maturity	1324
Linear + plateau	3.77
Tbase	5.20
DD anthesis	737
DD maturity	1324

Tplateau	34.5°C
Triangle Tbase, Topt, Tmax DD ant, DD mat	3.77 5.2, ∞, ? 737, 1324

Full crop models

Examine equations

This approach is in general much more difficult for crop models as a whole than for phenology models, since crop models involve multiple interacting equations. Therefore, it is important to develop other approaches to model inter-comparison.

Test multiple versions of a specific aspect of a crop model

This approach is analogous to that described for phenology models. One identifies a specific aspect of the crop models (for example, response to elevated CO₂ concentrations), and simulates using multiple formulations for that aspect, keeping the rest of the model unchanged. The differences in simulated values between the different formulations are then due just to differences in the aspect studied. Of course these differences are specific to the overall crop model in which the multiple formulations are embedded, to the specific contexts studied and to the specific parameter values used.

Several crop modelling platforms have been configured so that one can choose among multiple formulations for certain processes. For example, the STICS crop model (Brisson et al. 2003) allows one to add competition for assimilate between vegetative organs and reserve organs, to consider or not the geometry of the canopy when simulating radiation interception, to include or not water circulation in soil macropores, to choose among various descriptions of the root density profile, or to use a resistive approach to estimate the evaporative demand by plants rather than using a more empirical approach. Other crop modeling platforms also allow one to test multiple options for certain aspects of the model. For example, DSSAT (Jones et al. 2003) has two versions for simulating soil organic matter, among other options. APSIM is another modular modeling platform which allows easy testing of multiple versions of certain model functions (Keating et al. 2003).

More general modeling platforms, such as RECORD (Bergez et al. 2013), allow even more flexibility in choosing multiple versions of a model.

Compare models for contexts chosen to highlight a specific type of response

In this approach one compares different crop models, using a range of contexts which differ in some particular attribute, for example temperature or CO₂ concentration. Such tests highlight some specific aspect of the models. For example, (Asseng et al. 2014) used multiple models to simulate a range of contexts where the main factor of variability was temperature. This type of study does not specifically show which aspects of the models cause differences in simulated values, but it does pinpoint how simulated values differ when some particular environmental gradient is studied.

Use multiple outputs to separate different contributions to model variability

Crop models generally simulate a large number of output variables. This makes it possible to do multiple comparisons between models, in order to determine which outputs are similar and which have larger differences.

It is possible to go a step further, in order to identify which aspects of models are mostly responsible for differences in outputs. We illustrate with a simple example. One can write the equation for total evapotranspiration over the season (ET) as

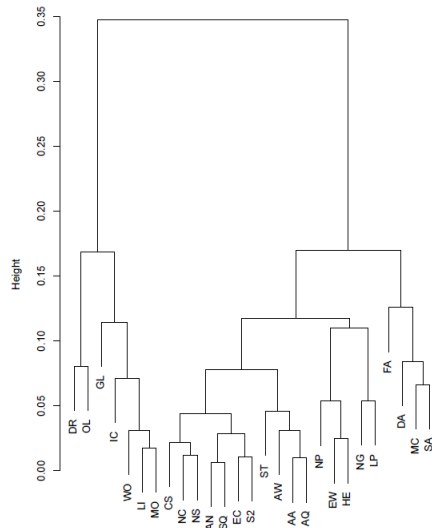
$$ET = ET_0 * (E_s/ET_0 + T_a/ET_0)$$

where ET_0 is potential evapotranspiration, E_s is soil evaporation and T_a is actual crop transpiration. When using a MME, one has the variance (over models) of both ET and of the terms on the right hand side of the equation. One can then evaluate how much of the overall variance in ET is due to differences in modeling ET_0 , and how much is due to differences in E_s/ET_0 or in T_a/ET_0 .

Identifying families of models

In this approach, one tries to identify families of models that have similar outputs. Once this classification is obtained, one can try to relate it to some underlying properties of the model, for example model complexity (as measured perhaps by number of parameters), or the way leaf area is modeled, using either a single leaf compartment or differentiating leaves by age, or some other criterion of model classification.

One approach to classifying models by similarity is to use a clustering algorithm, illustrated below for an ensemble of wheat models.



Conclusions

A major goal of crop model inter-comparison is model improvement, and an important intermediate step toward that goal is understanding in some detail how models differ. One approach here is to test multiple functions for the same process, and to analyse the differences engendered in simulated values. Another approach is to consider full crop models, but to simulate for a range of contexts which differ in some simple way (e.g. temperature gradient). A third approach is to use variance decomposition to relate model differences to specific aspects of the models. A fourth approach is to classify models by similarity in outputs, and try to relate that to model characteristics.

It is important to keep in mind that model differences are due not only to differences in functional form, but also to differences in parameterization. The latter may well be as important, or even more so, than the former.

Model inter-comparison, or comparison with data, refers to specific contexts. The conclusions may not be valid for other contexts.

References

- Asseng, S., et al. 2014. Rising temperatures reduce global wheat production. *Nat. Clim. Chang.* **in press**. doi: 10.1038/NCLIMATE2470.
- Bergez, J.-E., Chabrier, P., Gary, C., Jeuffroy, M.H., Makowski, D., Quesnel, G., Ramat, E., Raynal, H., Rousse, N., Wallach, D., Debaeke, P., Durand, P., Duru, M., Dury, J., Faverdin, P., Gascuel-Oudou, C., and Garcia, F. 2013. An open platform to build, evaluate and simulate integrated models of farming and agro-ecosystems. *Environ. Model. Softw.* **39**: 39–49. doi: 10.1016/j.envsoft.2012.03.011.
- Brisson, N., Gary, C., Justes, E., Roche, R., Mary, B., Ripoche, D., Zimmer, D., Sierra, J., Bertuzzi, P., Burger, P., Bussi re, F., Cabidoche, Y.M., Cellier, P., Debaeke, P., Gaudill re, J.P., H nault, C., Maraux, F., Seguin, B., and Sinoquet, H. 2003. An overview of the crop model STICS. *Eur. J. Agron.* **18**: 309–332.
- Jones, J.W., Hoogenboom, G., Porter, C.H., Boote, K.J., Batchelor, W.D., Hunt, L.A., Wilkens, P.W., Singh, U., Gijsman, A.J., and Ritchie, J.T. 2003. The DSSAT cropping system model. *Eur. J. Agron.* **18**: 235–265.
- Keating, B., Carberry, P., Hammer, G., Probert, M., Robertson, M., Holzworth, D., Huth, N., Hargreaves, J.N., Meinke, H., Hochman, Z., McLean, G., Verburg, K., Snow, V., Dimes, J., Silburn, M., Wang, E., Brown, S., Bristow, K., Asseng, S., Chapman, S., McCown, R., Freebairn, D., and Smith, C. 2003. An overview of APSIM, a model designed for farming systems simulation. *Eur. J. Agron.* **18**: 267–288. ELSEVIER SCIENCE BV, PO BOX 211, 1000 AE AMSTERDAM, NETHERLANDS. doi: 10.1016/S1161-0301(02)00108-9.
- Kumudini, S., Andrade, F.H., Boote, K.J., Brown, G.A., Dzotsi, K.A., Edmeades, G.O., Gocken, T., Goodwin, M., Halter, A.L., Hammer, G.L., Hatfield, J.L., Jones, J.W., Kemanian, A.R., Kim, S.-H., Kiniry, J., Lizaso, J.I., Nendel, C., Nielsen, R.L., Parent, B., St ckle, C.O., Tardieu, F., Thomison, P.R., Timlin, D.J., Vyn, T.J., Wallach, D., Yang, H.S., and Tollenaar, M. 2014. Predicting Maize Phenology: Intercomparison of Functions for Developmental Response to Temperature. *Agron. J.*: in press. doi: doi:10.2134/agronj14.0200; Published online 13 Aug. 2014.
- Ottman, M.J., Kimball, B.A., White, J.W., and Wall, G. 2012. Wheat growth 79 response to increased temperature from varied planting dates and 80 supplemental infrared heating. *Agron. J.* **104**: 716. doi: 10.2134/agronj2011.0212.