A taxonomy-based approach to shed light on the babel of mathematical models for rice simulation

Article *in* Environmental Modelling & Software · September 2016

DOI: 10.1016/j.envsoft.2016.09.007

Roberto Confalonieri
University of Milan

Myriam Adam
Cirad - La recherche agronomique pour le développement

Françoise Ruget
French National Institute for Agriculture, Food, and Environment (INRAE)

Tao Li
International Rice Research Institute

A taxonomy-based approach to shed light on the babel of mathematical models for rice simulation

Roberto Confalonieri[1*], Simone Bregaglio[1], Myriam Adam[2], Françoise Ruget[3], Tao Li[4, †], Toshihiro Hasegawa[5,‡], Xinyou Yin[6,‡], Yan Zhu[7,‡], Kenneth Boote[8,‡], Samuel Buis[3], Tamon Fumoto[5], Donald Gaydon[9], Tanguy Lafarge[2], Manuel Marcaida[4], Hiroshi Nakagawa[10], Alex C. Ruane[11], Balwinder-Singh[20], Upendra Singh[13], Liang Tang[7], Fulu Tao[14,21], Job Fugice[13], Hiroe Yoshida[10], Zhao Zhang[15], Lloyd T. Wilson[16], Jeff Baker[17], Yubin Yang[16], Yuji Masutomi[18], Daniel Wallach[19], Marco Acutis[1], Bas Bouman[4]

[1]University of Milan, Cassandra lab, Italy.
[2]CIRAD, UMR AGAP, F-34398 Montpellier, France.
[3]INRA, UMR1114 EMMAH, F-84914 Avignon, France.
[4]International Rice Research Institute, Los Baños, Philippines.
[5]National Institute for Agro-Environmental Sciences, Japan.
[6]Centre for Crop Systems Analysis, Wageningen University, the Netherlands.
[7]National Engineering and Technology Center for Information Agriculture, Jiangsu Key Laboratory for Information Agriculture, Jiangsu Collaborative Innovation Center for Modern Crop Production, Nanjing Agricultural University, China.
[8]University of Florida, Gainesville, Florida, USA.
[9]CSIRO Agriculture and Food, Brisbane, Australia.
[10]National Agriculture and Food Research Organization, Japan.
[11]NASA Goddard Institute for Space Studies, New York, USA.
[13]International Fertilizer Development Center, Muscle Shoals, Alabama, USA.
[14]Chinese Academy of Sciences, Institute of Geographical Sciences and Natural Resources Research, China.
[15]State Key Laboratory of Earth Surface Processes and Resource Ecology, Beijing Normal University, China.
[16]Texas A&M AgriLife Research & Extension Center, Beaumont, Texas, USA.
[17]United States Department of Agriculture, Agricultural Research Service, Big Spring, Texas, USA.
[18]Ibaraki University, College of Agriculture, Ibaraki-ken, Japan.
[19]INRA, UMR1248 Agrosystèmes et Développement Territorial, F-31 326 Castanet-Tolosan, France.
[20]CIMMYT, CG Block, NASC Complex, New Delhi-110012, India.
[21]Natural Resources Institute Finland (Luke), 01301 Vantaa, Finland.

[*]Corresponding Author, E-mail: roberto.confalonieri@unimi.it
[†]Leader of the AgMIP Rice Team. [‡]Members of the leading group of the AgMIP Rice Team.

Abstract

For most biophysical domains, differences in model structures are seldom quantified. Here, we used a taxonomy-based approach to characterise thirteen rice models. Classification keys and binary attributes for each key were identified, and models were categorised into five clusters using a binary similarity measure and the unweighted pair-group method with arithmetic mean. Principal component analysis was performed on model outputs at four sites. Results indicated that (i) differences in structure often resulted in similar predictions and (ii) similar structures can lead to large differences in model outputs. User subjectivity during calibration may have hidden expected relationships between model structure and behaviour. This explanation, if confirmed, highlights the need for shared protocols to reduce the degrees of freedom during calibration, and to limit, in turn, the risk that user subjectivity influences model performance.

Keywords

Model classification, model parameterisation, model ensemble, model structure, rice, uncertainty.

1. Introduction

"The sciences do not try to explain, they hardly even try to interpret, they mainly make models." This is one of the most famous of John von Neumann's quotes (von Neumann, 1955) and, despite its apparently provocative nature, presents a valuable perspective on how users or developers should view modelling. A mathematical model is a simplified representation of our knowledge about a real system, and of the way this knowledge is formalised (Malinvaud, 1964). During the second half of the 20th century, the crisis of the unitary reductionist conception of classical mathematical physics led scientists to gradually switch to a new paradigm in the way mathematics can be applied to the life sciences, largely based on the use of mathematical analogies (i.e., models) to describe real systems (Israel, 1993). A direct consequence of this transition is that mathematical models are by definition not unique (Israel, 1993), since none is capable of reproducing the underlying systems exactly, nature not being written in the language of mathematics, as per Galileo's thought.

The crop modelling community is increasingly conscious of the limitations of models, as demonstrated by the progressive abandonment of the expression *mechanistic models* in favour of *process-based models* (Dubitzky et al., 2013), which is a more suitable way to describe mathematical representations where empirical elements are used to describe processes that originate in cells and/or organelles but are observed at a higher hierarchical level (i.e., at crop canopy level; Acock and Acock, 1991). Our models are thus mathematical analogies useful for describing crop behaviour and derived from hypotheses for describing processes (e.g., photosynthesis or assimilate partitioning) that scientists verify using experimental data. Moreover, these models are strictly dependent on the context (agro-climatic, cultural, and social) in which such hypotheses are formulated, as well as on the specific objectives of their developers, which can range from improving our understanding of the interactions between plants and environment to supporting decision making (Passioura, 1996).

Of the models for rice simulation evaluated by Li et al. (2015) within the Rice Team of the Agricultural Model Intercomparison and Improvement Project (AgMIP, Rosenzweig et al., 2013), some (e.g., ORYZA2000, Bouman and van Laar, 2006) were originally proposed to formalise available knowledge and to draw attention to gaps in understanding (van Ittersum et al., 2003). Other models were developed targeting potential productivity (SIMRIW, Horie et al., 1995), management/decision support (CERES-Rice, Singh et al., 1993; H/H, Hasegawa and Horie, 1997; STICS, Brisson et al., 1998; APSIM, Holzworth et al., 2014), or ideotyping studies (GECROS, Yin and van Laar, 2005; GEMRICE, Yoshida et al., 2010, 2011; SAMARA, Kumar et al., 2016). Still others were designed to integrate knowledge from different domains, such as biogeochemistry (DNDC-Rice, Fumoto et al., 2008), microbiology (Gaydon et al., 2012a), and phytopathology (WARM, Confalonieri et al., 2009). To a certain extent, these different primary objectives reflect differences in the ways that processes are formalized in models.

The existence of differences in formalisation is further demonstrated by the (at least partial) coherence in the way we describe the states of domain in which crop dynamics are observed (e.g., aboveground biomass has a universally accepted meaning) compared to the range of meanings we give to model parameters to represent how crops behave within that domain. As an example, radiation use efficiency (RUE, biomass accumulated per unit solar radiation absorbed by the crop canopy) is a parameter included in five of the 13 models evaluated by Li et al. (2015). But the meaning of RUE varies among the models. In the CERES-Rice model, for instance, it is defined relative to total biomass, whereas in the WARM and STICS models it relates to aboveground biomass. In other models, such as SAMARA, radiation use efficiency is used to estimate gross photosynthesis, from which maintenance and growth respiration are subtracted. Thus, radiation use efficiency does not represent something belonging, in itself, to the real domain but is *just* a parameter of some of the possible representations we have constructed for what happens in that domain and, in turn, it has a precise meaning only within a specific model analogy. However, despite differences in definition and structure, crop models are expected to represent real systems with reasonable accuracy, and in most cases they actually do, thus – returning to von Neumann's perspective – justifying their existence and use in research and operational contexts.

Here, we analysed the relationships between the structures of mathematical analogies used for rice simulations and their outputs and, in the case of poor relationships, attempted to elucidate their causes. The work was made possible by a recent study performed within the activities of the AgMIP Rice Team (Li et al., 2015), where protocols for multi-model ensembles were applied.

2. Materials and methods

2.1. Rice models

The rice models used in this study were recently reviewed and evaluated by Li et al. (2015) and include both generic and rice-specific models. The generic models consisted of GECROS (Yin and van Laar, 2005), MCWLA (Tao and Zhang, 2013), and STICS (Brisson et al., 1998). These models, several of which have been used by other AgMIP teams (e.g., Asseng et al., 2013), were evaluated for different crop species (e.g., Biernath et al., 2011). The rice-specific models consisted of APSIM-Oryza (Gaydon et al., 2012b), CERES-Rice (Singh et al., 1993), DNDC-Rice (Fumoto et al., 2008), GEMRICE (Yoshida and Horie, 2010), H/H (Hasegawa and Horie, 1997), ORYZA2000 (Bouman and van Laar, 2006), RiceGrow (Tang et al., 2009), SIMRIW (Horie et al., 1995), SAMARA (Kumar et al., 2016), and WARM (Confalonieri et al., 2009). These models were either developed to focus on rice-based systems (e.g., GEMRICE, H/H, WARM, ORYZA2000) or, to varying extents, derived by modifying algorithms of existing models. In two cases, existing models for specific sub-domains (e.g., for rice growth and development) were included in simulation environments, facilitating the simulation of complex cropping and farming systems (APSIM-Oryza, DNDC-Rice). Some models share algorithms with others that belong to the same family (e.g., CERES-Rice shares approaches with CERES-Maize and CERES-Wheat).

Most of the models were developed in the late 1990s and 2000s, although a few were developed some years before (CERES-Rice, H/H, STICS); SIMRIW is the oldest of the models used in this study, having been initially developed in 1987, and later modified in 1993 and 2003 to increase its suitability for climate change studies. Major updates primarily consist of the incorporation of new approaches for specific sub-domains (GEMRICE, Yoshida et al., 2011; WARM, Confalonieri et al., 2012; STICS; Bergez et al., 2014) or extension to encompass processes not included in previous versions (APSIM-Oryza, Gaydon et al., 2012a). Other improvements are ongoing for many models, in part due to activities performed within the AgMIP Rice Team.

2.2. Analysis of model structure

Models were analysed using a three-step classification approach derived from taxonomic studies. The first step was to identify the main processes involved with growth and development; the second step was to define classification keys within each process, and the third step was to assign binary attributes for each classification key (Tables 1 to 3). As an example, within the process *Phenological development*, the classification key *Time step* has the attribute *hourly* that can be *true* or *false*. Within the same process, the classification key *Temperature at* has four attributes (*air, water, meristematic apex*, *canopy*) that can be either *true* or *false*. This led us to assigning a binary vector to each model.

In accordance with Wery (2005), Hay and Porter (2006), and Adam et al. (2010), six categories of processes were identified: phenological development, leaf area dynamics, biomass accumulation, partitioning to vegetative organs, and yield formation. Within each of these categories, a variable number of classification keys were identified. For example, *Partitioning to vegetative organs* has three classification keys: (i) *driven by*, (ii) *remobilisation*, and (iii) *temperature failure,* whereas other processes have as many as ten keys. *Biomass accumulation*, for instance, includes (i) *Time step*, (ii) *Temperature at* [air, water, meristematic apex, canopy], (iii) *Atmospheric* CO_2 *concentration* [considered or not], (iv) *Air relative humidity* [considered or not], (v) *Canopy representation*, (vi) *Photosynthesis* [RUE, leaf light response curve, Farquhar, N-driven], (vii) *Saturation to light of enzymatic chains* [considered or not], (viii) *Atmospheric* CO_2 *concentration effects on respiration* [considered or not], (ix) *Respiration* [explicit or implicit], and (x) *Organ senescence*. The classification keys can be modified for use in other modelling studies because they depend on the models used by the AgMIP Rice Team and on the targeted conditions. For example, nitrogen-limiting conditions are not considered in this study; thus, related processes were not considered for the classification and, for models that include algorithms for nitrogen, they were either switched off or the model was run simulating fully fertilised conditions. Classification keys can help to identify (i) modelling approaches, such as those used for photosynthesis; (ii) driving variables, such as temperature (air, water, meristematic apex, canopy) driving phenological development; (iii) response functions (e.g., linear or not); and (iv) presence/absence of algorithms for sub-processes, such as photoperiod. For this reason, we decided to use the generic term *attribute* for the different choices available for each classification key. Note that attributes can refer to processes that are reproduced using the same conceptual approach, although the way knowledge is formalised for each approach can be modified in different ways for each model. For instance, in the case of a non-linear temperature response function for phenological development, equations may differ but still have similar shapes.

Table 1. Analysis of model structure: phenological development and leaf area dynamics.

Process	**Classification key**	**Attribute**	APSIM-Oryza	CERES-Rice	DNDC-Rice	GECROS	GEMRICE	MCWLA	H/H	ORYZA2000	RiceGrow	SAMARA	SIMRIW	STICS	WARM
Phenological development	Time step	- hourly		●		●					●				●
	Temperature at	- air	●	●	●	●	●	●	●	●	●	●	●		
		- water		●					●						
		- meristematic apex													●
		- canopy												●	
	Response to temperature	- linear	●	●				●		●		●		●	
	Atmospheric CO_2 concentration	- considered					●						●		
	Limitation for high temperature	- considered	●	●	●	●	●	●		●	●	●	●	●	●
	Photoperiod	- considered	●	●		●	●	●	●	●	●	●	●	●	●
	Response to photoperiod	- linear	●	●		●		●		●				●	●
	Temperature failure	- yes						●							
Leaf area dynamics	Temperature at	- air	●	●	●		●	●	●	●	●	●	●		
		- water			●				●						
		- canopy												●	●
	Green leaf area increase	- C-driven	●	●		●	●			●	●				●
		- N-driven		●	●	●	●		●						
		- organogenesis		●								●			
		- temperature-driven	●	●	●			●		●			●	●	
	Leaf senescence	- C-driven		●					●		●				
		- N-driven		●		●	●		●						
		- organogenesis										●			
		- temperature-driven	●	●				●		●				●	●
		- age structure			●				●				●		
	Self-shading	- considered		●		●		●	●	●	●	●			
Reference[§]			1	2	3	4	5	6	7	8	9	10	11	12	13

§ 1: Gaydon et al. (2012b); 2: Singh et al. (1993); 3: Fumoto et al. (2008); 4: Yin and van Laar (2005); 5: Yoshida and Horie (2010); 6: Tao and Zhang (2013); 7: Hasegawa and Horie (1997); 8: Bouman and van Laar (2006); 9: Tang et al. (2009); 10. Kumar et al. (2016); 11: Horie et al. (1995); 12: Brisson et al. (1998); 13: Confalonieri et al. (2009).

Table 2. Analysis of model structure: biomass accumulation.

Process	Classification key	Attribute	APSIM-Oryza	CERES-Rice	DNDC-Rice	GECROS	GEMRICE	MCWLA	H/H	ORYZA2000	RiceGrow	SAMARA	SIMRIW	STICS	WARM
Biomass accumulation	Time step	- hourly				●	●		●						●
	Temperature at	- air	●	●	●		●	●	●	●	●	●	●		
		- canopy				●								●	●
	Atmospheric CO_2 concentration	- considered	●	●	●	●	●	●	●	●	●		●	●	●
	Air relative humidity	- considered			●	●		●	●						
	Canopy representation	- multi-layer	●			●	●		●	●	●				
	Photosynthesis	- radiation use efficiency		●								●	●	●	●
		- leaf light response curve	●				●			●	●				
		- Farquhar			●	●		●	●						
		- N-driven		●		●	●		●						
	Saturation to light of enzymatic chains	- considered	●		●				●	●	●			●	●
	Atm. CO_2 concentration on respiration	- considered						●							
	Respiration	- explicit (gross photosynthesis)	●		●	●	●	●	●	●	●	●			
		- implicit (net photosynthesis)		●									●	●	●
	Biomass senescence	- C-driven				●					●				
		- N-driven		●		●									
		- development stage-driven	●	●	●			●	●	●			●		
		- temperature-driven		●					●			●		●	
Reference§			1	2	3	4	5	6	7	8	9	10	11	12	13

§ 1: Gaydon et al. (2012b); 2: Singh et al. (1993); 3: Fumoto et al. (2008); 4: Yin and van Laar (2005); 5: Yoshida and Horie (2010); 6: Tao and Zhang (2013); 7: Hasegawa and Horie (1997); 8: Bouman and van Laar (2006); 9: Tang et al. (2009); 10. Kumar et al. (2016); 11: Horie et al. (1995); 12: Brisson et al. (1998); 13: Confalonieri et al. (2009).

Table 3. Analysis of model structure: partitioning to vegetative organs and yield formation.

Process	**Classification key**	**Attribute**	APSIM-Oryza	CERES-Rice	DNDC-Rice	GECROS	GEMRICE	MCWLA	H/H	ORYZA2000	RiceGrow	SAMARA	SIMRIW	STICS	WARM
Partitioning to vegetative organs	Driven by	- supply/demand		●	●	●	●					●			
		- development stage	●	●			●	●	●	●	●			●	●
	Remobilization	- considered	●	●		●	●			●				●	●
	Temperature failure	- considered						●							
Yield formation	Atmospheric CO_2 concentration	- considered				●	●	●					●		
	Grain number	- simulated	●	●		●	●			●				●	
	Panicle number	- simulated		●											
	Spikelet sterility	- heat	●	●			●	●		●	●	●	●		●
		- cold	●	●			●	●	●	●	●	●	●		●
	Yield determination	- grain growth rate		●	●	●	●					●			
		- development stage-driven	●	●						●					●
		- dynamic harvest index						●	●		●		●	●	
Reference§			1	2	3	4	5	6	7	8	9	10	11	12	13

§ 1: Gaydon et al. (2012b); 2: Singh et al. (1993); 3: Fumoto et al. (2008); 4: Yin and van Laar (2005); 5: Yoshida and Horie (2010); 6: Tao and Zhang (2013); 7: Hasegawa and Horie (1997); 8: Bouman and van Laar (2006); 9: Tang et al. (2009); 10. Kumar et al. (2016); 11: Horie et al. (1995); 12: Brisson et al. (1998); 13: Confalonieri et al. (2009).

Our approach for the analysis of model structure constrains attributes to binary values (0 [false] or 1 [true]). Most of the models analysed in this study simulate processes not listed in Tables 1–3. We focused on the processes involved with rice growth and development under conditions of unlimited water and nitrogen, and in the absence of effects of pathogens, insects, or weeds.

We classified the models based on the Sokal and Sneath I (1973) binary similarity measure (Choi et al., 2010). Assuming $X = (x_1, \ldots, x_i, \ldots, x_n)$ and $Y = (y_1, \ldots, y_i, \ldots, y_n)$ as two vectors of n attributes, with each attribute assigned binary values, the Sokal and Sneath I similarity coefficient (S; unitless, 0–1) is calculated according to Eq. 1:

$$S = \frac{a}{a + 2b + 2c} \tag{1}$$

where a is the number of attributes x_i and y_i equal to 1 for both vectors; b is the number of attributes for which x_i is 0 and y_i is 1; and c is the number of attributes for which x_i is 1 and y_i is 0. This similarity coefficient excludes attributes for which both x_i and y_i are 0 (negative matches), because the possible lack of attributes in the vectors could prevent negative matches to refer to similarity between X and Y (Sokal and Sneath, 1963). A similarity coefficient was then used to perform an agglomerative hierarchical clustering (Cormack, 1971) using the unweighted pair-group method with arithmetic mean (UPGMA; Sneath and Sokal, 1973). The threshold for model clusters definition was based on the largest decrease in Shannon's entropy between a node and the next one. In practice, the metric $\frac{1}{entropy[node(i-1)-entropy[node(i)]]}$ was minimised in order to identify the truncation value.

2.3. Analysis of the differences in model structures compared to those in model outputs

Results of the model classification procedure were compared to differences in the results generated by the models in the study conducted by Li et al. (2015). In particular, we used simulated results, after model calibration, for the experiments performed under non-limiting conditions for water and nitrogen (Table 4).

Table 4. Datasets used for analysing differences in model outputs. Of the experiments presented in Li et al. (2015), only those characterised by non-limiting conditions for water, nutrients, weeds, pests, and diseases were used.

Site	**Institute**	**Latitude N**	**Longitude E**	**Climate**	**Year**	**Transplanting date**	**Cultivar name**	**Ecotype**	**Growth duration**
Los Baños, Philippines	IRRI[1]	14°6′	121°9′	Tropical, humid	1985	February 5	IR58	Indica	120-125 days
					1986	February 6			
Ludhiana, India	PAU[2], IRRI	30°54′	75°48′	Subtropical, semi-arid	2008	July 6	PAU201	Indica	130-135 days
					2009	July 6			
Nanjing, China	NAU[3]	32°56′	118°59′	Subtropical, semi-humid	2007	June 29	Wuxiangjing 14	Japonica	ca 150 days
Shizukuishi, Japan	NIAES, NARO/TART[4]	39°41′	140°57′	Cool temperate, humid	1998	May 22	Akitakomachi	Japonica	ca 150 days
					1999	May 20			
					2000	May 23			

[1] International Rice Research Institute, Los Baños, Philippines.

[2] Punjab Agricultural University, Ludhiana, India.

[3] Nanjing Agricultural University, Nanjing, China.

4 National Institute for Agro-Environmental Sciences, Tsukuba, Japan; NARO/TART = National Agricultural Research Organization/Tohoku Agricultural Research Center.

The experiments cover a wide variety of environments in which rice is grown in Asia. Climates range from tropical humid (Los Baños, Philippines) to cold temperate humid (Shizukuishi, Japan). During the rice growing season, the highest temperatures were from Ludhiana (India), with mean maximum and minimum air daily temperature exceeding 33.5 °C and 24.5 °C, respectively, whereas sub-optimal temperatures were measured for the three Shizukuishi experiments (with mean maximum and minimum daily air temperatures of 26.1 °C and 15.9 °C, respectively), and intermediate thermal conditions characterised the rice seasons in Los Baños and Nanjing (China). Most of the variation in the thermal regimes occurred during the vegetative phase. Further experimental details are available in Table 4 and in Li et al. (2015).

For each experiment, simulated growth duration (GD; days), aboveground biomass (AGB; Mg ha^{-1}) and panicle biomass (PB) at physiological maturity (Mg ha^{-1}), and maximum leaf area index (around flowering; LAI; m^2 m^{-2}) were analysed. The results of each site were processed using principal component analysis (PCA) to account for correlations among variables. In particular, PCA was performed for each study site to determine the effects of site × year interactions on model behaviour. Distances among models in the principal component spaces were analysed in light of each models' classification within the clusters, as described in the previous section.

3. Results

3.1. Model structure

The UPGMA analysis identified CERES-Rice (cluster BI) as differing the most from the other models (Fig. 1). Three model clusters were found each composed of two models: AI, consisting of GECROS and GEMRICE; AII, consisting of DNDC-Rice and H/H; BII, consisting of STICS and WARM. The largest cluster (AIII) consisted of MCWLA, RiceGrow, APSIM-Oryza, ORYZA2000, Samara, and SIMRIW.

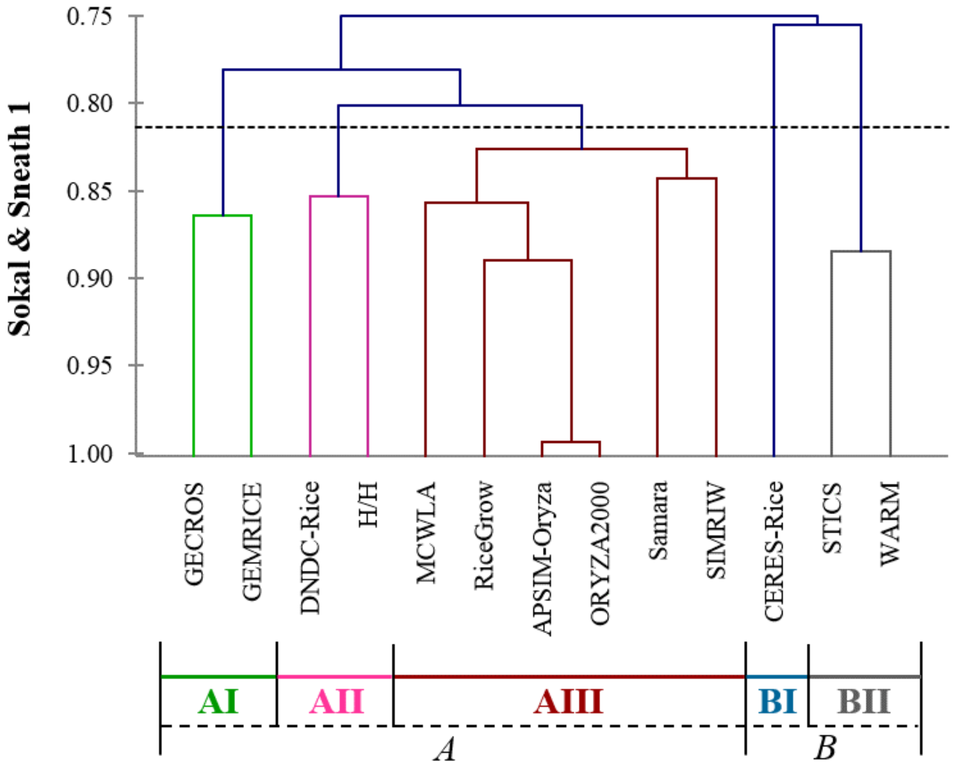

Figure 1. Dendrogram showing model clusters obtained using the classification criteria presented in Tables 1–3. Agglomerative hierarchical clustering was performed using the Sokal and Sneath I similarity coefficient and the unweighted pair-group method with arithmetic mean.

The differences between CERES-Rice and the other models largely reflected CERES-Rice providing more detailed description of the processes involved in the dynamics of plant organs during the crop cycle (senescence, remobilization, and yield formation). The distinctiveness of GECROS and GEMRICE (AI) was due to their explicit focus on genotype × environment interaction, and thus on trait descriptions and on feedback mechanisms. STICS and WARM (cluster BII) share a net photosynthesis approach based on the concept of radiation use efficiency, albeit modified to allow the simulation of processes normally considered in more complex approaches (e.g., saturation of enzymatic chains). The inclusion of DNDC-Rice and H/H in the same cluster (AII) was due to the H/H module for crop growth being used as partial substitute for SUCROS (van Keulen et al., 1982) algorithms within the current version of DNDC-Rice. Cluster AIII is the largest and most heterogeneous of the clusters. Models belonging to this cluster lack a module for energy balance and use air temperature to drive phenological development, leaf area dynamics, and photosynthesis. Three models in this cluster (RiceGrow, APSIM-Oryza, and ORYZA2000) implement SUCROS-type algorithms for canopy representation and photosynthesis, but differ for other processes.

In general, the main differences among the models concerned partitioning of assimilates to vegetative organs and yield formation, whereas fewer or more minor differences were observed for leaf area dynamics and biomass accumulation and, most notably, for phenological development.

3.2. Relationship between model structures and simulation outputs

The first two PCA components explained over 75% of the total variance in simulated output, with about 85% of the variance explained for the simulations in Los Baños and Shizukuishi sites. For the same sites, component 1 included information mostly related to PB, AGB and LAI, whereas GD was predominant in component 2. These sites had low values of simulated LAI, PB, and AGB, apparently due to the lower radiation during the vegetative phase compared to the other two sites. The way in which components 1 and 2 represented information from the four variables analysed for the Ludhiana and Nanjing sites was less distinct, with LAI related to both components. In general, component 1 explained about half of the total variance (from 45.3% in Ludhiana to 59.2% in Shizukuishi), whereas the second component explained about 25% of the variance, with the exception of Ludhiana, for which 32% of the total variance was explained. Although some authors have suggested to perform PCA with at least three items for each factor (e.g., Costello and Osborne, 2005), the PCA we performed had just descriptive objectives. Our aim was indeed to provide readers with a synthetic representation of outputs for the different sites, years, and models. However, the meaningfulness of the analysis is confirmed by the eigenvalues for the two principal components, which were always greater than one (e.g., Kaiser, 1960; Fabrigar et al., 1999).

Results of the PCAs for the four experimental sites differed markedly (Fig. 2), underscoring a clear effect of site on the relative differences among the models. The within-site year effect was considerably less important (data not shown).

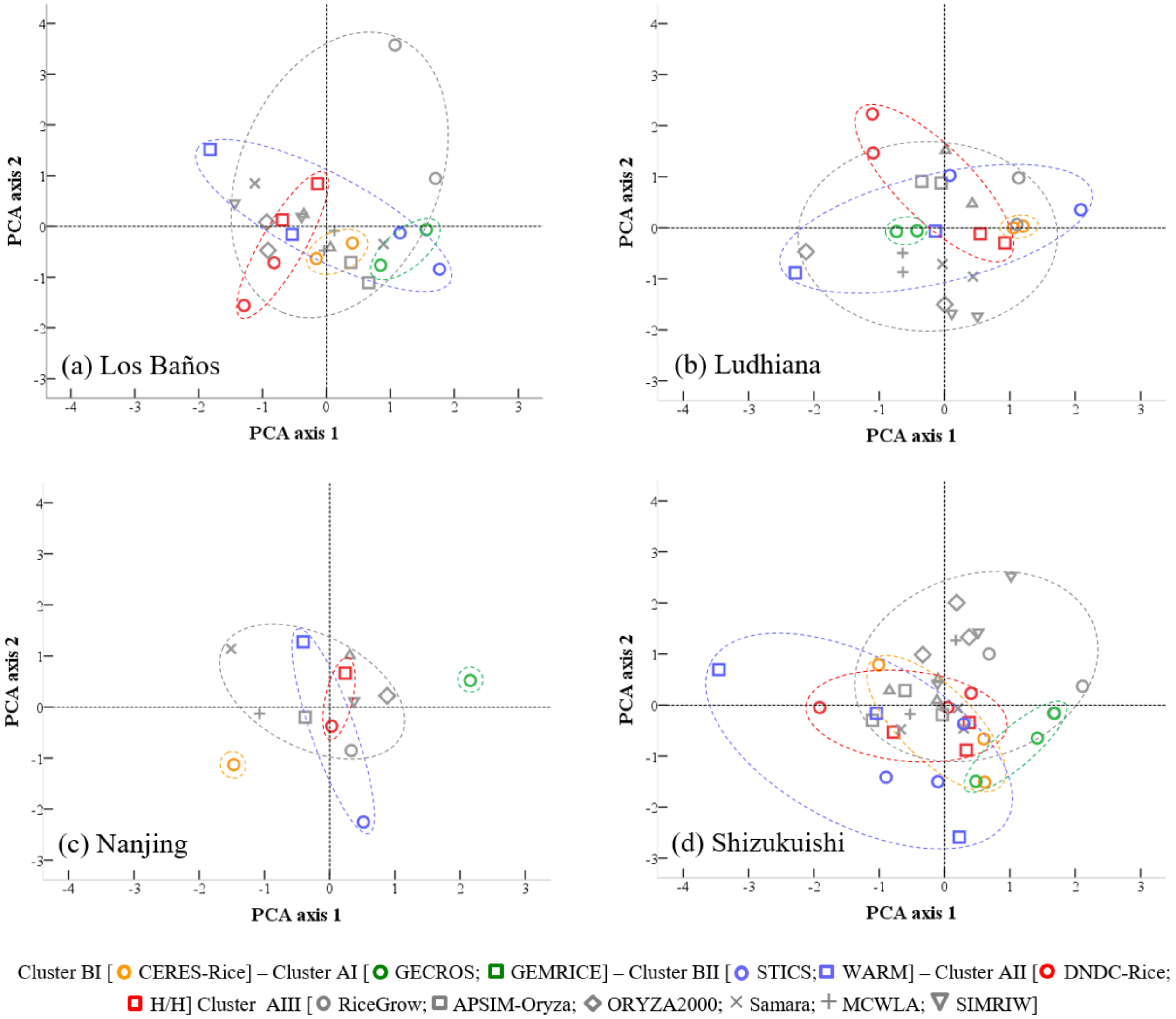

Figure 2. Principal component analysis (PCA) of simulated growth duration, maximum leaf area index (around flowering), aboveground biomass and panicle biomass at maturity; PC scores are shown.

The results of the UPGMA performed on model structure did not reflect the relative differences among model outputs, with clear patterns in model results rarely identifiable (Fig. 2). Cluster AII (including DNDC-Rice and H/H) was the only cluster that displayed a consistent trend in output among the sites. In particular, for simulations of Los Baños and Shizukuishi, the distances of the points due to model and year effects were comparable (Figs. 2.a and 2.d). However, for the Ludhiana site, H/H results were explained almost entirely by component 1, whereas DNDC-Rice results were equally explained by component 1 (inversely compared to H/H) and component 2. To a certain extent, cluster AIII exhibited a certain coherence for the model × year interaction in the coldest site (Shizukuishi; Fig. 2.d), although this pattern did not hold for Los Baños and Ludhiana (Figs. 2.a and 2.b). In particular, ORYZA2000 and APSIM-Oryza largely shared the same structure (Fig. 1) and were also similar in terms of output for Los Baños (Fig. 2.a), although larger differences in output existed for Nanjing (Fig. 2.c) and Shizukuishi (Fig. 2.d). In some cases, the year effect generated larger differences than those attributable to the models, even within the same cluster. Examples of this behaviour include GEMRICE in Los Baños, ORYZA2000 in Ludhiana, and WARM in Shizukuishi.

Figure 3.a demonstrates how different outputs can be produced (in this case starting from the beginning of the heading stage at Los Baños in 1985) by two models (the two versions of Oryza) that are very similar according to the cluster analysis presented in Fig. 1 (cluster AIII). On the other hand, AGB curves that almost overlapped were simulated by two models (CERES-Rice and WARM) belonging to distinct clusters (BI and BII, respectively). For Shizukuishi 1999 (Fig. 3.b), the simulation results were very similar for CERES-Rice and WARM (at least until the ripening stage), whereas differences in the ORYZA2000 and APSIM-Oryza outputs were, as expected, negligible. Figure 3.c illustrates another example, in which the two versions of Oryza generated different AGB values, whereas two other models that were very different in structure (STICS and MCWLA, belonging to clusters BII and AIII, respectively) provided very similar outputs. As expected, Fig. 3.d presents distinct AGB values from models belonging to different clusters (BII for WARM and STICS, BI for CERES-Rice and AIII for MCWLA).

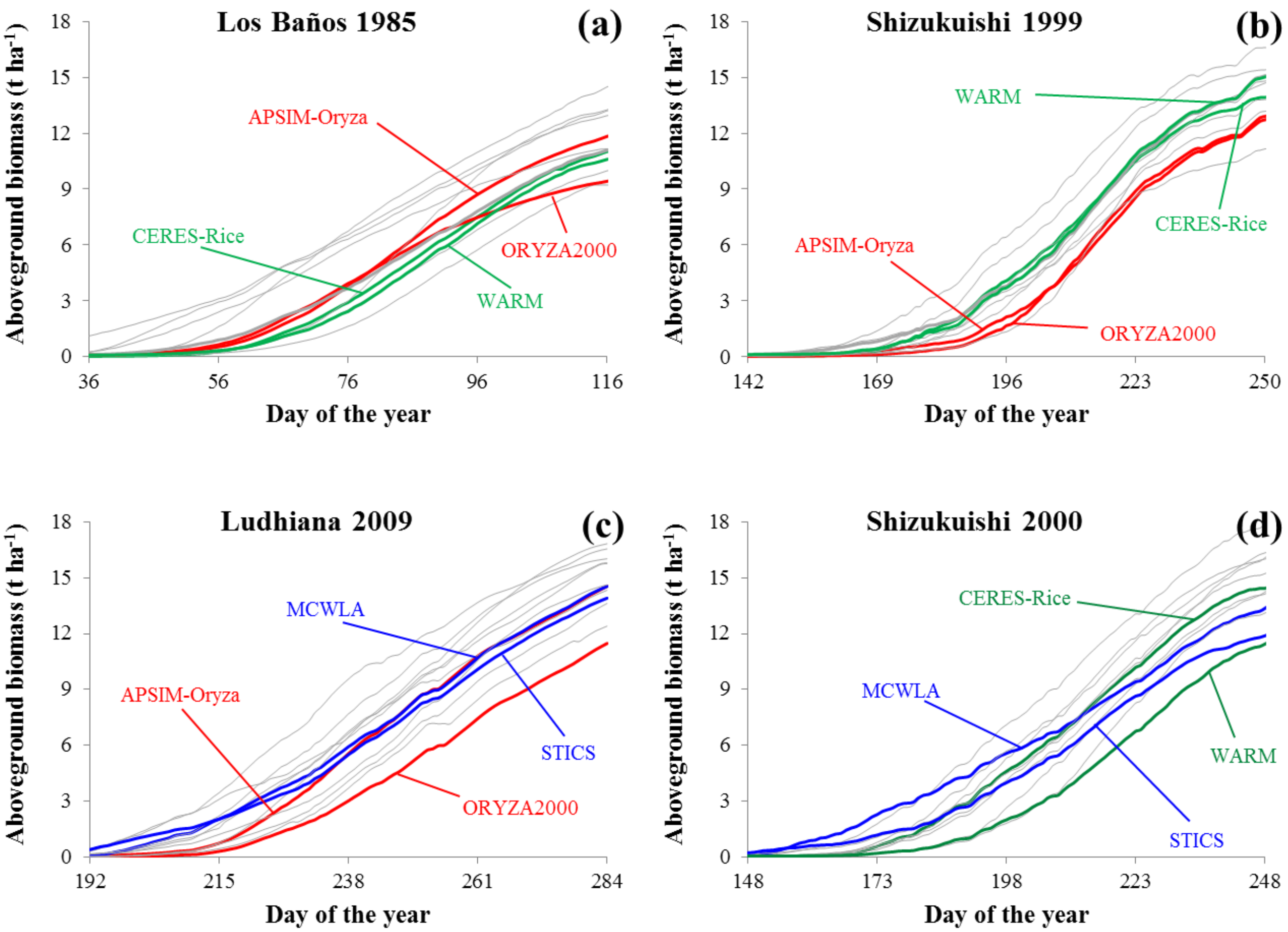

Figure 3. Sample results underlying differences among aboveground biomass (AGB) data simulated by the different models. Among the models highlighted, APSIM-Oryza, ORYZA2000, and MCWLA belong to cluster AIII, WARM and STICS to cluster BI1, and CERES-Rice to cluster BI. Los Baños 1985, Shizukuishi 1999 and 2000, and Ludhiana 2009 refer to the datasets presented in Table 4.

4. Discussion

Attempts were recently made to cluster crop models based on relationships between simulated yields and weather variables during the growing season (Ruane et al., 2016). However, our study is the first to classify crop models based on their structure (i.e., on how knowledge about different processes is formalised) with a transparent and reproducible methodology. Analysis of the structure of the 13 models included in our study resulted in five clusters, the largest composed of six models. Even within a cluster, however, there were large differences in model complexity. For instance, the largest cluster included four models based on gross photosynthesis: MCWLA based on a Farquhar-type approach (Farquhar et al., 1980), the others (RiceGrow and the two versions of Oryza) based on leaf light response curve (De Wit, 1978). In contrast, in the same cluster, Samara and SIMRIW incorporate a simpler radiation use efficiency-based net photosynthesis approach (Warren Wilson, 1967). The same consideration of models having different degrees of complexity within the same cluster is also true for yield formation: the largest cluster includes one model based on grain growth rate (Samara), two models partitioning assimilates to grains as a function of development stage (the two versions of Oryza), and three models based on harvest index to compute grain yield from aboveground biomass (MCWLA, RiceGrow, SIMRIW). The independence of clusters from model complexity is valid for other clusters as well, albeit with such differences being less evident given that, with the exception of cluster BI, each of the other clusters includes only two models.

Differences in model outputs within each cluster were comparable with those among clusters (e.g., GEMRICE in Los Baños, WARM in Shizukuishi; Fig. 2), thus casting a shadow on the existence of strong relationships between model structure and model outputs. Similar conclusions were reported by Ruane et al. (2016), who analysed 27 wheat models. Although the authors of that study clustered models based solely on the relationships between outputs and seasonal weather (and not structure), they acknowledged that the common heritage of different modelling approaches – which should, theoretically, lead to structural analogies – was not reflected clearly in model behaviour. This suggests that factors other than model structure influenced simulation results, and that part of the uncertainty in model outputs is not directly attributable to the models themselves (Wallach, 2011; Rötter et al., 2012). Our results revealed model behaviours that were, to an extent, unexpected. On the one hand, models that *might have been expected* to provide different results because of pronounced differences in their structure actually converged in terms of output. On the other hand, models that would be expected to provide similar results because of their structure can generate markedly different outputs. The first case might be explained by considering the approaches used for some sub-processes (e.g., photosynthesis) as different ways to formalise our knowledge about the real system. Indeed, given that mathematical analogies are not “true” by definition, they instead represent alternative approximations, with some performing in a very coherent way under certain environmental and management conditions but not necessarily under others. Examples of such behaviour are illustrated in Figs. 3.a and 3.b, with CERES-Rice (cluster BI; Fig. 1) and WARM (cluster BII) providing aboveground biomass values that almost completely overlapped for the Los Baños 1985 and Shizukuishi 1999 datasets, despite major differences in the approaches used for most processes (Tables 1–3) and the different environmental conditions under which rice was grown in the two datasets. The same consideration is clearly supported by the agreement between the MCWLA (cluster AIII) and STICS (cluster BII) outputs for the Ludhiana 2009 dataset (Fig. 3.c).

The second case (unexpected divergence of outputs generated by models similar in structure) may be explained by user subjectivity during calibration (Janssen and Heuberger, 1995), which can affect model behaviour (Sinclair and Seligman, 2000) to such an extent that models with similar structures may behave differently. Indeed, because not all models contain automated calibration frameworks – which must themselves be parameterised and thus are, to an extent, also exposed to users' subjectivity – calibration in the crop modelling community is often still performed via trial-and-error, although such calibration method is increasingly being abandoned in other modelling fields (van Vliet et al., 2016). The impact of user subjectivity during calibration was recently quantified by Confalonieri et al. (2016), who arranged five crop models (levels of the factor *model*) and 20 users in a randomised block design with four replicates. They discovered that differences among the outputs of different models were significant in 30% of the maize datasets only, given the large variability in the outputs of each model when calibrated by different users. The situation is even more complex given that user subjectivity can lead to relevant differences in the outputs of the same model in the absence of any biophysical incoherence in the values of calibrated parameters. The clearest example of models providing different outputs despite their structural coherence is that of ORYZA2000 and APSIM-Oryza (Figs. 3.a and 3.c). For some datasets, the two models simulated distinctly different aboveground biomass during the reproductive phase (e.g., Los Baños 1985) or the whole growth cycle (Ludhiana 2009). Differences in aboveground biomass at maturity were as high as 30%; however, this was partly due to the differences in the soil modules implemented in the two versions of Oryza. Indeed, although we selected only datasets with the highest nitrogen application rates, minor differences in nitrogen availability were simulated by the two models.

The last consideration suggests algorithms for different processes can differentially impact the output of each, and thus classification keys and related attributes (Tables 1 to 3) should be weighted in a different manner before cluster analyses are performed. One way of minimizing subjectivity when assigning weights (that might affect model structure analysis) is to define a hierarchy among the processes based on the relevance of the parameters related to each attribute. In such a case, a variety of sensitivity analysis techniques are available that can be used to quantify the relevance of input parameters on model outputs (Jakeman et al., 2006), for which comparisons were made for WARM, one of the crop models used in this study (Confalonieri et al., 2010). However, two issues prevent the assignment of weights to attributes based on sensitivity analysis results. The first is involved with the different number of parameters used within the algorithms for each attribute in each model. For each of the sub-processes corresponding to each classification key/attribute, indeed, models differ in the number of parameters used. Moreover, when a process is not explicitly included in a specific model, related attributes will not be weighted at all, and deriving weights for classification keys and attributes in a different way for each model is impractical. The second problem is involved with the effect of explored conditions on sensitivity analysis results. The relevance of model parameters on simulation results can vary greatly under different climates and management scenarios, although this tendency differs among models (Confalonieri et al., 2012).

5. Conclusions

Our classification method was effective in clustering models in a transparent and reproducible way. However, the analysis of the relationship between model structures and model outputs (PCA performed on outputs for each experimental site) revealed that differences in structure can produce similar outputs, and coherence in approaches can lead to large differences in simulation results. Among the explanations for these behaviours – somehow observed also by Ruane et al. (2016) for wheat models – are user subjectivity during the calibration as well as the history of model parameterizations before the study performed by Li et al. (2015), given that initial parameter values can affect calibration results. If confirmed in further studies, these explanations should prompt the modelling community to develop rigorous standard protocols designed to restrict the degrees of freedom during calibration, and to limit, in turn, the risk that user subjectivity will alter model functioning. These protocols can be based, for example, on multi-user calibration (Diekkrüger et al., 1995; Confalonieri et al., 2016) or on automated optimisation algorithms (Muleta and Nicklow, 2005) and, in any case, by an *a priori* analysis of model behaviour via sensitivity analysis techniques (Tarantola and Saltelli, 2003). Of course, in the case of crop modelling, the situation is complicated by large differences in model complexity (in terms of the number of model parameters) and by the presence of parameters with the same name but different meanings in different models.

Indeed, although different authors have proposed good practices for model use over recent decades (Sinclair and Seligman, 2000; Jakeman et al., 2006; Wallach et al., 2006), the crop modelling community does not appear to be paying sufficient attention to uncertainty produced by model users.

An improvement to the methodology used to cluster models could derive from the assignment of weights to the classification keys and attributes based on their differential impacts on model outputs. However, despite sensitivity analysis theoretically being a suitable technique, no solution was found because of the presence of a different number of parameters in the models' algorithms for the same attribute, and the known variability in sensitivity analysis results while changing conditions of applications prevent its use for classification purposes.

Authorship

This paper represents the combined effort of thirty researchers who appear as co-authors. We comprise the AgMIP-Rice project team. RC coordinated the study and wrote the first draft of the paper. SB, MA, RC, FR motivated the study and developed the first version of data analysis, later improved by TL, TH, KB, DG, LTW, JB, YY, DW. All authors provided crop model simulations and description/explanations of the characteristics of the models. All authors contributed to the discussion of results.

Acknowledgements

The authors are grateful to the AgMIP project, which have made the collaboration behind this study possible. FT is supported by the National Science Foundation of China (31561143003) and the Academy of Finland, PLUMES project (277403). TH's participation was supported by MAFF and the Global Environment Research (S-10-2) of MOE, Japan. YZ's work was supported by the National High-Tech Research and Development Program of China (2013AA100404). RC was partly supported by EC-FP7 under Grant Agreement No. 613817

(MODEXTREME). Nanjing Agricultural University, International Rice Research Institute, Ministry of Agriculture, Forestry and Fisheries and Ministry of the Environment of Japan, and Cassandra lab. of the University of Milan supported the annual research planning meetings of the AgMIP Rice Team. USDA is an equal opportunity provider and employer.

References

Acock, B., Acock, M.C., 1991. Potential for using long-term field research data to develop and validate crop simulators. Agron. J. 83, 56-61.

Adam, M., Ewert, F., Leffelaar, P.A., Corbeels, M., van Keulen, H., Wery, J., 2010. CROSPAL, software that uses agronomic expert knowledge to assist modules selection for crop growth simulation. Environ. Model. Soft. 25, 946-955.

Asseng, S., Ewert, F., Rosenzweig, C., Jones, J.W., Hatfield, J.L., Ruane, A.C., Boote, K.J., Thorburn, P.J., Rötter, R.P., Cammarano, D., Brisson, N., Basso, B., Martre, P., Aggarwal, P.K., Angulo, C., Bertuzzi, P., Biernath, C., Challinor, A.J., Doltra, J., Gayler, S., Goldberg, R., Grant, R., Heng, L., Hooker, J., Hunt, L.A., Ingwersen, J., Izaurralde, R.C., Kersebaum, K.C., Müller, C., Naresh Kumar, S., Nendel, C., O'Leary, G., Olesen, J.E., Osborne, T.M., Palosuo, T., Priesack, E., Ripoche, D., Semenov, M.A., Shcherbak, I., Steduto, P., Stöckle, C.O., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Travasso, M., Waha, K., Wallach, D., White, J.W., Williams, J.R., Wolf, J., 2013. Uncertainty in simulating wheat yields under climate change. Nat. Clim. Change 3, 827-832.

Bergez, J.E., Raynal, H., Launay, M., Beaudoin, N., Casellas, E., Caubel, J., Chabrier, P., Coucheney, E., Dury, J., Garcia de Cortazar-Atauri, I., Justes, E., Mary, B., Ripoche, D., Ruget, F., 2014. Evolution of the STICS crop model to tackle new environmental issues: New formalisms and integration in the modelling and simulation platform RECORD. Environ. Modell. Soft. 62, 370-384.

Biernath, C., Gayler, S., Bittner, S., Klein, C., Högy, P., Fangmeier, A., Priesack, E., 2011. Evaluating the ability of four crop models to predict different environmental impacts on spring wheat grown in open-top chambers. Eur. J. Agron. 35, 71-82.

Bouman, B.A.M., van Laar, H.H., 2006. Description and evaluation of the rice growth model ORYZA2000 under nitrogen-limited conditions. Agr. Syst. 87, 249-273.

Brisson, N., Mary, B., Ripoche, D., Jeuffroy, M.E., Ruget, F., Nicollaud, B., Gate, P., Devienne-Barret, F., Antonioletti, R., Durr, C., Richard, G., Beaudoin, N., Recous, S., Tayot, X., Plenet, D., Cellier, P., Machet, J.-M., Meynard, J.M., Delécolle, R., 1998. STICS: a generic model for the simulation of crops and their water and nitrogen balances. I. Theory and parameterization applied to wheat and corn. Agronomie 18, 311-346.

Choi, S.-S., Cha, S.-H., Tappert, C.C., 2010. A survey of binary similarity and distance measures. J. Syst. Cybern. Informatics 8, 43-48.

Confalonieri, R., Bellocchi, G., Bregaglio, S., Donatelli, M., Acutis, M., 2010. Comparison of sensitivity analysis techniques: a case study with the rice model WARM. Ecol. Model. 221, 1897-1906.

Confalonieri, R., Bregaglio, S., Acutis, M., 2012. Quantifying plasticity in simulation models. Ecol. Model. 225, 159-166.

Confalonieri, R., Orlando, F., Paleari, L., Stella, T., Gilardelli, C., Movedi, E., Pagani, V., Cappelli, G., Vertemara, A., Alberti, L., Alberti, P., Atanassiu, S., Bonaiti, M., Cappelletti, G., Ceruti, M., Confalonieri, A., Corgatelli, G., Corti, P., Dell'Oro, M., Ghidoni, A., Lamarta, A., Maghini, A., Mambretti, M., Manchia, A., Massoni, G., Mutti, P., Pariani, S., Pasini, D., Pesenti, A., Pizzamiglio, G., Ravasio, A., Rea, A., Santorsola, D., Serafini, G., Slavazza, M., Acutis, M., 2016. Uncertainty in crop model predictions: What is the role of users? Environ. Modell. Softw. 81, 165-173.

Confalonieri, R., Rosenmund, A.S., Baruth, B., 2009. An improved model to simulate rice yield. Agron. Sustain. Dev. 29, 463-474.

Cormack, R.M., 1971. A review of classification. J. Roy. Stat. Soc. A Sta. 134, 321-367.

Costello, A.B., Osborne, J.W., 2005. Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. Prac. Assessment Rese. Eval. 10, 1-9.

De Wit, C.T., 1978. Simulation of assimilation, respiration and transpiration of crops. Pudoc, Wageningen, The Netherlands.

Diekkrüger, B., Söndgerath, D., Kersebaum, K.C., McVoy, C.W., 1995. Validity of agroecosystem models. A comparison of results of different models applied to the same data set. Ecol. Model. 81, 3-29.

Dubitzky, W., Wolkenhauer, O., Cho, K.-H., Yokota, H., 2013. Encyclopedia of Systems Biology. Springer, New York, NY, pp. 2367.

Fabrigar, L.R., Wegener, D.T., MacCallum, R.C., Strahan, E.J., 1999. Evaluating the use of exploratory factor analysis in psychological research. Psychol. Methods 3, 272-299.

Farquhar, G.D., von Caemmerer, S., Berry, J.A., 1980. A biochemical model of photosynthesis CO_2 assimilation in leaves of C3 species. Planta 149, 78-90.

Fernández, A., Gómez, S., 2008. Solving non-uniqueness in agglomerative hierarchical clustering using multidendrograms. J. Classif. 25, 43-65.

Fumoto, T., Kobayashi, K., Li, C., Yagi, K., Hasegawa, T., 2008. Revising a process-based biogeochemistry model (DNDC) to simulate methane emission from rice paddy fields under various residue management and fertilizer regimes. Glob. Change Biol. 14, 382-402.

Gaydon, D.S., Probert, M.E., Buresh, R.J., Meinke, H., Timsina, J., 2012a. Capturing the role of algae in rice crop production and soil organic carbon maintenance. Eur. J. Agron. 39, 35-43.

Gaydon, D.S., Probert, M.E., Buresh, R.J., et al., 2012b. Rice in cropping systems: modelling transitions between flooded and non-flooded soil environments. Eur. J. Agron. 39, 9-24.

Hasegawa, T., Horie, T., 1997. Modelling the effect of nitrogen on rice growth and development. In: Applications of systems approaches at the field level, Vol. 2 (Eds. Kropff, M.J. et al.), pp. 243-257. London, UK: Kluwer Academic Publishers.

Hay, R.K.M., Porter, J.R., 2006. The Physiology of Crop Yield, second ed. Blackwell Publishing, Ames, Iowa, USA.

Holzworth, D.P., Huth, N.I., deVoil, P.G., Zurcher, E.J., Herrmann, N.I., McLean, G., Chenu, K., van Oosterom, E., Snow, V., Murphy, C., Moore, A.D., Brown, H., Whish, J.P.M., Verrall, S., Fainges, J., Bell, L.W., Peake, A.S., Poulton, P.L., Hochman, Z., Thorburn, P.J., Gaydon, D.S., Dalgliesh, N.P., Rodriguez, D., Cox, H., Chapman, S., Doherty, A., Teixeira, E., Sharp, J., Cichota, R., Vogeler, I., Li, F.Y., Wang, E., Hammer, G.L., Robertson, M.J., Dimes, J., Carberry, P.S., Hargreaves, J.N.G., MacLeod, N., McDonald, C., Harsdorf, J., Wedgwood, S., Keating, B.A., 2014. APSIM - Evolution towards a new generation of agricultural systems simulation. Environmental Modelling & Software, 62, 327-350.

Horie, T., Nakagawa, H., Centeno, H.G.S., Kropff, M.J., 1995. The rice crop simulation model SIMRIW and its testing. In: Modeling the impact of climate change on rice production in Asia (Eds Matthews RB, Kropff MJ, Bachelet D, van Laar HH), pp. 51-56. Wallingford, Oxon, UK: CAB International in association with International Rice Research Institute.

Israel, G., 1993. The emergence of biomathematics and the case of population dynamics. A revival of mechanical reductionism and Darwinism. Sci. Context 6, 469-509.

Jakeman, A.J., Letcher, R.A., Norton, J.P., 2006. Ten iterative steps in development and evaluation of environmental models. Environ. Model. Softw. 21, 602-614.

Janssen, P.H.M., Heuberger, P.S.C., 1995. Calibration of process-oriented models. Ecol. Model. 83, 55-66.

Kaiser, H.F., 1960. The application of electronic computers to factor analysis. Educ. Psychol. Meas. 20, 141-151.

Kumar, U., Laza, M.R., Soulié, J.C., Pasco, R., Mendez, K.V.S., Dingkuhn, M., 2016. Compensatory phenotypic plasticity in irrigated rice: Sequential formation of yield components and simulation with SAMARA model, Field Crop. Res. 193, 164-177.

Li, T., Hasegawa, T., Yin, X., Zhu, Y., Boote, K., Adam, M., Bregaglio, S., Buis, S., Confalonieri, R., Fumoto, T., Gaydon, D., Marcaida, M. III, Nakagawa, H., Oriol, P., Ruane, A.C., Ruget, F., Singh, B., Singh, U., Tang, L., Tao, F., Wilkens, P., Yoshida, H., Zhang, Z., Bouman, B., 2015. Uncertainties in predicting rice yield by current crop models under a wide range of climatic conditions. Glob. Change Biol. 21, 1328-1341.

Malinvaud, E., 1964. Méthodes statistiques de l'économétrie. Dunod, Paris, 634 pp.

Muleta, M.K., Nicklow, J.W., 2005. Sensitivity and uncertainty analysis coupled with automatic calibration for a distributed watershed model. J. Hydrol. 306, 127-145.

Passioura, J.B., 1996. Simulation models: science, snake oil, education, or engineering? Agron. J. 88, 690-694.

Rosenzweig, C., Jones, J.W., Hatfield, J.L., Ruane, A.C., Boote, K.J., Thorburn, P., Antle, J.M., Nelson, G.C., Porter, C., Janssen, S., Asseng, S., Basso, B., Ewert, F., Wallach, D., Baigorria, G., Winter, J.M., 2013. The

Agricultural Model Intercomparison and Improvement Project (AgMIP): Protocols and pilot studies. Agr. Forest. Meteorol. 170, 166-182.

Rötter, R.P., Palosuo, T., Kersebaum, K.C., Angulo, C., Bindi, M., Ewert, F., Ferrise, R., Hlavinka, P., Moriondo, M., Nendel, C., Olesen, J.E., Patil, R.H., Ruget, F., Takáč, J., Trnka, M., 2012. Simulation of spring barley yield in different climatic zones of Northern and Central Europe: A comparison of nine crop models. Field Crop. Res. 133, 23-36.

Ruane, A.C., Hudson, N.I., Asseng, S., Camarano, D., Ewert, F., Martre, P., Boote, K.J., Thorburn, P.J., Aggarwal, P.K., Angulo, C., Basso, B., Bertuzzi, P., Biernath, C., Brisson, N., Challinor, A.J., Doltra, J., Gayler, S., Goldberg, R., Grant, R.F., Heng, L., Hooker, J., Hunt, L.A., Ingwersen, J., Izaurralde, R.C., Kersebaum, K.C., Kumar, S.N., Müller, C., Nendel, C., O'Leary, G., Olesen, J.E., Osborne, T.M., Palosuo, T., Priesack, E., Ripoche, D., Rötter, R.P., Semenov, M.A., Shcherbak, I., Steduto, P., Stöckle, C.O., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Travasso, M., Waha, K., Wallach, D., White, J.W., Wolf, J., 2016. Multi-wheat-model ensemble responses to interannual climate variability. Environ. Modell. Soft. 81, 86-101.

Sinclair, T.R., Seligman, N., 2000. Criteria for publishing papers on crop modelling. Field Crop. Res. 68, 165-172.

Singh, U., Ritchie, J.T., Godwin, D.C., 1993. A Users' Guide to CERES-Rice V2.10, Simulation manual. 131 pp. IFDC, Muscle Shoals, AL, USA.

Sokal, R.R., Sneath, P.H.A., 1963. Principles of numerical taxonomy. W.H. Freeman & Co., San Francisco, 359 pp.

Sneath, P.H.A., Sokal, R.R., 1973. Numerical taxonomy: the principles and practice of numerical classification. W.H. Freeman & Co., San Francisco, 573 pp.

Tang, L., Zhu, Y., Hannaway, D., Meng, Y., Liu, L., Chen, L., Cao, W., 2009. RiceGrow: a rice growth and productivity model. NJAS - Wageningen Journal of Life Sciences, 57, 83-92.

Tao, F., Zhang, Z., 2013. Climate change, high temperature stress, rice productivity and water use in eastern China: a new super-ensemble-based probabilistic projection. J. Appl. Meteorol. Clim. 52, 531-551.

Tarantola, S., Saltelli, A., 2003. SAMO 2001: methodological advances and innovative applications of sensitivity analysis. Reliab. Eng. Syst. Safety 79, 121-122.

Van Ittersum, M.K., Leffelaar, P.A., Van Keulen, H., Kropff, M.J., Bastiaans, L., Goudriaan, J., 2003. On approaches and applications of the Wageningen crop models. Eur. J. Agron. 18, 201-234.

Van Keulen, H., Penning de Vries, F.W.T., Drees, E.M., 1982. A summary model for crop growth. In: Simulation of Plant Growth and Crop Production. PUDOC, Wageningen, the Netherlands, pp. 87-98.

Van Vliet, J., Bregt, A.K., Brown, D.G., van Delden, H., Heckbert, S., Verburg, P.H., 2016. A review of current calibration and validation practices in land-change modeling. Environ. Modell. Soft. 82, 174-182.

Von Neumann, J., 1955. Method in the Physical Sciences. In: The Unity of Knowledge (Ed. Leary, L.G.). Doubleday, Garden City, NY, 306 pp.

Wallach, D., 2011. Crop model calibration: a statistical perspective. Agron. J. 103, 1144-1151.

Wallach, D., Makowski, D., Jones, J.W., 2006. Working with Dynamic Crop Models – Evaluation, Analysis, Parameterisation, and Applications. Elsevier B.V., Amsterdam, The Netherlands. 447 pp.

Warren Wilson, J., 1967. Ecological data on dry-matter production by plants and plant communities. In: Bradley, E.F., Denmead, O.T. (Eds.) The collection and processing of field data. Interscience, New York, pp. 77-123.

Wery, J., 2005. Differential effects of soil water deficit on the basic plant functions and their significance to analyse crop responses to water deficit in indeterminate plants. Aust. J. Agr. Res. 56, 1201-1209.

Yin, X., van Laar, H.H., 2005. Crop Systems Dynamics: an ecophysiological simulation model for genotype-by-environment interactions. 151 pp. Wageningen, The Netherlands: Wageningen Academic Publishers.

Yoshida, H., Horie, T., 2010. A model for simulating plant N accumulation, growth and yield of diverse rice genotypes grown under different soil and climatic conditions. Field Crop. Res. 117, 122-130.

Yoshida, H., Horie, T., Nakazono, K., Ohno, H., Nakagawa, H., 2011. Simulation of the effects of genotype and N availability on rice growth and yield response to an elevated atmospheric CO_2 concentration. Field Crop. Res. 124, 433-440.