

# Incorporating realistic trait physiology into crop growth models to support genetic improvement

K. J. Boote<sup>\*</sup>, J. W. Jones and G. Hoogenboom<sup>\*</sup>

Agricultural and Biological Engineering Department, University of Florida, Gainesville, FL 32611, USA

<sup>\*</sup>Corresponding author's e-mail address: [kjboote@ufl.edu](mailto:kjboote@ufl.edu)

Guest Editor: Karine Chenu

Editor-in-Chief: Stephen P. Long

**Citation:** Boote KJ, Jones JW, Hoogenboom G. 2021. Incorporating realistic trait physiology into crop growth models to support genetic improvement. *In Silico Plants* **2021**: diab002; doi: 10.1093/insilicoplants/diab002

## ABSTRACT

*In silico* plant modelling is the use of dynamic crop simulation models to evaluate hypothetical plant traits (phenology, processes and plant architecture) that will enhance crop growth and yield for a defined target environment and crop management (weather, soils, limited resource). To be useful for genetic improvement, crop models must realistically simulate the principles of crop physiology responses to the environment and the principles by which genetic variation affects the dynamic crop carbon, water and nutrient processes. Ideally, crop models should have sufficient physiological detail of processes to incorporate the genetic effects on these processes to allow for robust simulations of response outcomes in different environments. Yield, biomass, harvest index, flowering date and maturity are emergent outcomes of many interacting genes and processes rather than being primary traits directly driven by singular genetics. Examples will be given for several grain legumes, using the CSM-CROPGRO model, to illustrate emergent outcomes simulated as a result of single and multiple combinations of genotype-specific parameters and to illustrate genotype by environment interactions that may occur in different target environments. Specific genetically influenced traits can result in  $G \times E$  interactions on crop growth and yield outcomes as affected by available water,  $CO_2$  concentration, temperature, and other factors. An emergent outcome from a given genetic trait may increase yield in one environment but have little or negative effect in another environment. Much work is needed to link genetic effects to the physiological processes for *in silico* modelling applications, especially for plant breeding under future climate change.

**KEYWORDS:** Climate change; CROPGRO-model; crop simulation; crop physiology; genotype-specific-parameters (GSP); genotype by environment interaction; genetic improvement; phenotype.

## 1. INTRODUCTION

Genetic improvement of crops is important for increasing world food supply, in view of the continued increase in world population on a fixed arable land base with limited water and fertilizer resources (Hickey *et al.* 2019). Enhancing crop yield through science-based breeding has occurred over many decades, a task that has been further accelerated in recent years through molecular technologies using DNA-based markers (Varshney *et al.* 2020). Plant breeders, for many years, have attempted, and succeeded in many cases, to model plant ideotypes that result in higher yields (Donald 1968) starting in the 1960s with the shorter semi-dwarf rice (*Oryza sativa*) cultivars that did not lodge under increased N supply (Chandler 1969). With improved tools (molecular genetics, more advanced crop models) and an urgent mission of adapting to climate change and feeding an increased world

population, it is timely to use these tools to hypothesize genetic improvement in yield. With the recent advancements in dynamic crop growth simulation (Thorburn *et al.* 2018; Boote 2019), crop models have excellent potential for analysing past genetic improvement from experimental data and for proposing plant ideotypes for target environments (Suriharn *et al.* 2007, 2011; Peng *et al.* 2008; Putto *et al.* 2008). Crop simulation models have potential for creating 'virtual crop cultivars' and for assisting the breeder's selection criteria as well as integrating molecular marker-based information (White and Hoogenboom 2003; Hoogenboom *et al.* 2004; Chenu *et al.* 2009; Hammer *et al.* 2010, 2016; Muller and Martre 2019; Oliveira *et al.* 2021), and for genetic enhancement of important traits that contribute to yield improvement in different target environments (Hammer *et al.* 2004; Hammer and Jordan 2007; Technow *et al.* 2015; Yin and Struik 2016).

Using crop simulation models to evaluate traits for genetic improvement is not new and has been attempted a number of times during the past 40 years (Duncan et al. 1978; Landivar et al. 1983; Elwell et al. 1987; Boote and Tollenaar 1994; Hammer et al. 1996; Boote et al. 2001, 2003; Chapman et al. 2003). What is different is that molecular genetics information was not available during those early efforts and connection to true genes was not feasible. What is perhaps unique about those early efforts is that those models were relatively mechanistic for their time. Elwell et al. (1987) used the SOYMOD Soybean model (Curry et al. 1975), which included leaf-level photosynthesis and mass flow phloem translocation. Boote and Tollenaar (1994) used an early version of CROPGRO that had leaf-to-canopy assimilation, explicit tissue compositions, growth respiration and single seed growth dynamics. Also during that time, there were attempts at genetic variation with simpler models, although the interpretation of genetic information to alter radiation use efficiency (RUE), for example, is difficult because RUE is the emergent outcome of the expression of many genes. However, even models that used emergent traits as dynamic state variables have been successful for simulating genetic variability influences on life cycle phase durations, and are useful for defining best cultivar life cycles for given target environments (Löffler et al. 2005; Technow et al. 2015).

In the past two decades (2000–20), crop modellers have attempted to connect genetic yield improvement more closely to genes and quantitative trait loci (QTLs). White and Hoogenboom (1996) were among the first to add specific gene effects into a crop model for dry bean (*Phaseolus vulgaris*). Other examples include QTL-based predictors of specific leaf area of rice (Yin et al. 1999), leaf area expansion in maize (*Zea mays*) (Tardieu 2003), time to flower in rice (Yin et al. 2005a, b), time to heading in wheat (*Triticum*) (Zheng et al. 2013) and time to flower, pod-set and maturity of soybean (*Glycine max*) (Messina et al. 2006). Quantitative trait locus connection to reproductive life cycle progression appears to be convenient and relatively easy to achieve. Life cycle is important to fit cultivars to their target environment, and as Messina et al. (2006) showed, those QTL effects on life cycle account for a considerable amount of cultivar yield variability.

Others have tried to connect QTLs and genes in crop models to simulate more complex phenotypic outcomes for breeding purposes (Yin et al. 1999, 2004; Chapman et al. 2003; Hammer et al. 2004, 2005, 2010, 2016; Cooper et al. 2005; Chenu et al. 2009). For example, Chenu et al. (2009) linked 11 QTL markers to three genotype-specific parameters (GSPs) affecting leaf elongation rate (LER) and one GSP affecting the anthesis-to-silking interval (ASI) of maize. The LER-connected coefficients were temperature sensitivity, vapor pressure deficit (VPD)-sensitivity and water-potential sensitivity. These sensitivities for leaf area expansion and grain-set (latter via ASI) were placed into a modified Agricultural Production Systems Simulator (APSIM)-Maize model, and all possible re-combinations of the QTLs were simulated to evaluate maize yield under drought conditions. The re-combinations were coded with the same approach as Messina et al. (2006) to indicate presence (1) or absence (0) of a given allele. Also important in this case, is that these features of hourly leaf elongation sensitivity to temperature, VPD and water potential were placed into the APSIM model (Hammer et al. 2010) which already responds to management (sowing date, sowing density, irrigation, N fertilization, soil characteristics), and environmental conditions (temperature,

irradiance, rainfall), such that the APSIM model generated emergent phenotypic outcomes for simulated drought conditions. They simulated a number of G × E interactions resulting from combinations of the QTL markers for maize in different drought environments.

Early crop simulation models were designed to be specific for a given crop, with little if any emphasis on genetic differences within a species (see BACROS and SUCROS examples in Penning de Vries and van Laar 1982). Other models such as CERES-Maize and CERES-Wheat (Ritchie et al. 1998) have cultivar differences but with relatively few cultivar-specific parameters (GSPs). Both types of models are generic in the sense of simulating crop C, N and water balances, and crop life cycle progression using generally accepted principles of biology and biogeochemical processes but with minimal consideration for real genetics. In these cases, there was little linkage to the true genetics (e.g. DNA, molecular markers). Most recent crop models are designed to simulate cultivars within a species; nevertheless, those models' GSPs are artificial constructs (parameters) that reproduce different phenotypic life cycles, daylength sensitivities, productivities and seed size/growth rate traits without considering molecular genetics information. Although there is a need to link crop models to molecular genetics, fully accomplishing this task will take decades due to the complexity and lack of sufficient details at the process level that are normally captured by the crop simulation models. The goal in this paper is to describe how GSPs in current crop simulation models can be used to best consider traits for use in genetic improvement. Using crop simulation models to evaluate genetic improvement has advantages over pure bioinformatics analysis. This is because the models have embedded in them the functional sensitivities of physiological processes to temperature, water deficit, N stress and other stresses, and they incorporate C, N and water balances. The latter features are missing in pure bioinformatics approaches; thus, yield predictions that integrate dynamic crop models in whole-genome prediction are more predictive of responses to weather, soil and management (Technow et al. 2015; Cooper et al. 2016; Messina et al. 2018). Weather, soil and management components are very important in order to fully understand the dynamic interactions between genetics and environmental conditions (Tsuji et al. 1998). We will briefly review general principles of crop simulation models and then introduce ways in which the crop models handle cultivar differences and genetic input.

## 2. DYNAMIC CROP SIMULATION MODELS—PROCESSES AND RESPONSES TO ENVIRONMENT

Dynamic crop growth models compute the crop carbon (C) balance on a daily (or hourly) basis, based on rates of photosynthesis, C losses to growth and maintenance respiration, C losses to senescence/abscission of plant parts and partitioning of the net C to produce dry matter of different plant organs (Boote et al. 1998). Mathematical relationships for process sensitivities to environmental variations are included in the models. Those equations, rules and sensitivities of processes to weather and environment have been learned from analyses of measurements and studies of crop physiologists.

In addition to simulating soil N and C balance processes associated with organic matter decomposition and mineral N dynamics, the models also simulate the crop N balance on a daily time

step, with N input from root N uptake and/or  $N_2$ -fixation, N losses to senescence/abscission, N partitioning to different organs and N mobilization from one organ to another (Boote *et al.* 2009). Likewise, there are equations and sensitivities of root N uptake,  $N_2$ -fixation, cost of N reduction and N mobilization and partitioning that are parameterized from measurements. The daily crop-soil N balance simulates that tissue N concentrations will vary over time and, therefore, affect other processes such as photosynthesis, partitioning and grain growth.

The soil water balance is also extremely important in crop models as most crops are grown under rainfed conditions. An important principle that links the crop C gain to the water balance is the obligate stomatal coupling of  $CO_2$  uptake (photosynthesis) to water vapour escape (transpiration) from the stomatal pores. The processes affecting water balance include inputs from rainfall and irrigation considering infiltration versus run-off, water movement through the soil (leaching and lateral flow), as well as evaporative losses from the soil surface and the crop transpiring surfaces (Ritchie 1998; Boote *et al.* 2008). The soil-crop water balance usually honours principles of conservation of water and energy, on a daily or shorter time step, using evapotranspiration methods such as the Penman–Monteith evapotranspiration equations (Allen *et al.* 1998).

Crop development is a fourth type of process in the models that leads to crop growth stage progression and crop phenology, which are very important in all dynamic crop growth models. Simply stated, crop development is driven by growing degree days (GDD) or photo-thermal time unit accumulation for successive phases of the crop from sowing to maturity, that allows for the crop to fit in a given growing season niche.

The word ‘dynamic’ is very important in the context of crop growth models. This means that the crop responds dynamically to each day’s weather and soil conditions, as influenced by that day’s already existing crop state (leaf area index (LAI), amount of biomass, root length density, etc.). That differs considerably from statistical regression models. Some models actually simulate and scale up from leaf-to-canopy assimilation on an hourly time step (Boote and Pickering 1994; Pickering *et al.* 1995; Grant *et al.* 1995) although Class A weather data (solar radiation,  $T_{max}$ ,  $T_{min}$  and rainfall) are typically only available on a daily basis.

Finally, crop models integrate over time these multiple process rates with sensitivities to environment, and they simultaneously honour C, N and soil water balances as the crops progress through their life cycle sequences. Thus, crop models represent inherent resource limitations and environmental sensitivities and illustrate that yield requires inputs for water, N and other nutrients, solar radiation,  $CO_2$  and temperature over time during growing seasons.

### 3. GENETICS AND GSPS IN CROP SIMULATION MODELS

The above-described process-oriented models are relatively generic in their structure and simulation of vegetative and reproductive development, photosynthesis, respiration, translocation, partitioning and reproductive growth processes. However, the desire to accurately simulate the growth and development of particular genotypes within a given species resulted in the incorporation of cultivar-specific information in the modelling structure as done with

the Decision Support System for Agrotechnology (DSSAT) crop models beginning in DSSAT V3.5 (Boote *et al.* 1998; Tsuji *et al.* 1998) and subsequent versions (Boote *et al.* 2003; Jones *et al.* 2003; Hoogenboom *et al.* 2019a, b).

White and Hoogenboom (2003) proposed six levels by which genetic details are included in crop simulation models:

1. Generic model with no reference to species.
2. Species-specific model with no reference to genotypes.
3. Genetic differences represented by cultivar-specific parameters.
4. Genetic differences represented by specific alleles, with gene action represented through linear effects on model parameters.
5. Genetic differences represented by genotypes, with gene action explicitly simulated based on knowledge of regulation of gene expression and effects of gene products.
6. Genetic differences represented by genotypes, with gene action simulated at the level of interactions of regulators, gene products and other metabolites (in other words, the full genetic architecture relative to current crop state and its environment).

Most crop simulation models originated at level 2 for a given crop, based on experience with growth measurements and the developer’s knowledge of a candidate crop. The early models developed in the ‘School of de Wit’ were at level 1, including BACROS and SUCROS (Bouman *et al.* 1996). Some models remain at level 2, but many are now at level 3, with genetic differences represented by cultivar-specific parameters (GSPs). No candidate models have approached levels 5 and 6, while the GeneGro models for common bean (White and Hoogenboom 1996; Hoogenboom *et al.* 1997; Hoogenboom and White 2003) and soybean (Messina *et al.* 2006) and for wheat (White *et al.* 2008) have attempted level 4, with actions of specific genes (allelic 0 vs. 1). Zheng *et al.* (2013) developed a gene-based model for time to heading in wheat that was incorporated into APSIM-Wheat, based on vernalization genes (Vrn-A1, Vrn-B1, Vrn-D1) and a photoperiod sensitivity (Ppd-D1) gene. The predictability of heading was 4.3 days for 210 spring wheat lines over many sowing dates and locations in Australia. Recently, Hwang *et al.* (2017) developed a simple crop phenology model based on QTL marker information. Oliveira *et al.* (2021) used a similar approach to integrate a dynamic QTL-based module (Vallejos, CE., personal communication, 2020) into the CSM-CROPGRO-Drybean model for predicting first flower appearance.

For most crop models at level 3, there are two types of genetic information in the models that allow them to simulate unique growth and development responses of specific genotypes under specific environments. First, they have crop parameters and relationships (and this represents genetics as well) that are hard-wired in the computer code or inputs from species files that create the crop species type (maize or soybean or wheat). Species traits and relationships are defined in the computer code for CERES-Maize along with a small read-in species file (Ritchie *et al.* 1998). By contrast, CROPGRO has generic source code that allows the same code to be used for many species (up to 18 at present), but the species parameters and relationships are specified in a read-in species file. In addition, there are GSPs in read-in cultivar files that distinguish differences among cultivars,



varieties or hybrids within a species. While model developers may make this distinction for convenience of model users, in reality this is all genetically controlled and there should be no real distinction between species traits and cultivar traits within a species. For example, cardinal temperature parameterization of processes is defined in the CROPGRO species file at present, but cultivars may actually differ, for example, in heat tolerance of grain-set (see later example). It is a situation of initially being parsimonious with cultivar GSPs, especially when information is lacking.

In this paper, we will present examples with the DSSAT crop models, particularly the CROPGRO-Legume model (Boote et al. 1998; Hoogenboom et al. 2019b) using V4.7 release available from [www.DSSAT.net](http://www.DSSAT.net), to describe genetic traits and how they are simulated to evaluate genetic yield improvement for various target environments. Genotype-specific parameters for the CROPGRO model are defined in Table 1. The most important GSPs are those that define daylength sensitivities and phase durations (EM-FL, FL-SH, FL-SD, SD-PM) that affect crop life cycle (Table 1). For example, soybean is daylength-sensitive (critical short daylength (CSDL) = 12.58 h and PPSEN = 0.311 for MG 5 cultivar), whereas peanut (*Arachis hypogaea*) is not daylength-sensitive (PPSEN = 0.000). For simulating phenology and life cycle of winter cereals, vernalization parameters are important in addition to daylength and phase-duration parameters. For CERES-Maize, phase durations are described by GDD, while for CROPGRO, photothermal days (ptd) are used. CROPGRO has six vegetative GSPs affecting early leaf size (SIZFL), specific leaf area under optimum conditions (SLAVR), duration of leaf appearance (FL-VS), rate of leaf appearance (TRIFOL), duration of leaf expansion (FL-LF) beyond flowering and leaf photosynthesis (LFMAX). Reproductive GSPs affect potential seed size (WTPSD), seeds per pod (SDPDV), individual grain-growth rate (set by 1/SFDUR), the ptd duration of pod addition (PODUR) and threshing percentage (THRSH). Seed protein (SDPRO) and lipid (SDLIP) compositions are also defined as cultivar GSPs; these have become more important with the emphasis on nutrition security, especially under climate change.

Three example species within CROPGRO in Table 1 have dramatic differences in life cycle, especially in the duration of the seed-filling phase (SD-PM), with dry bean having the shortest and peanut having the longest seed-filling phase (Table 1; Fig. 1). Not surprisingly, the yield potential of the three crops generally follows that trend, with a lower yield for the shorter cycle dry bean crop. There are also differences in potential yield due to seed composition, with oil and protein requiring more energy for synthesis compared to carbohydrates. Dry bean has a short life cycle and lower yield, but compensates by having more rapid LAI development along with higher sowing density of 25 plants per m<sup>2</sup> compared to 18 plants per m<sup>2</sup> for soybean and 16 plants per m<sup>2</sup> for peanut (Fig. 1A). Peanut, under good fungicide treatment, does not self-senesce leaves, while dry bean and soybean have a rapid senescence of LAI that is rapid after the beginning maturity stage (first mature pods). Peanut is an example of an indeterminate crop, with a long slow phase of pod addition (PODUR), less than 1.00 partitioning intensity (XFRT) and a long period of seed-filling (SFDUR and SD-PM).

The CERES-Maize model has six GSPs (Ritchie et al. 1998) shown in Table 2. There are three GSPs related to life cycle and daylength sensitivity (P1, P2 and P5, which describe crop life cycle progression). P1

is GDD from emergence to end-juvenile, while P2 is daylength sensitivity which affects an internal phase to end of floral initiation. The outcome of P1 and P2 affects an internal phase that determines time from floral initiation to flowering. There are two reproductive parameters, G2 which is genetic potential number of grains per plant, and G3 which is single kernel growth rate under optimum conditions. PHINT is an additional GSP, which is meant to represent the rate of leaf appearance (PHINT being the GDD per leaf tip appearance). However, PHINT also re-scales simulated LAI (with a larger PHINT value reducing LAI). There are additional coefficients for potential leaf size and leaf longevity for the IXIM-Maize model, which is somewhat more mechanistic than CERES-Maize. The IXIM-Maize model has leaf-to-canopy photosynthesis (Lizaso et al. 2005) and more detailed relations for leaf area growth (Lizaso et al. 2003). Table 2 illustrates that longer life cycle hybrids grown in Iowa produce higher biomass and grain yield at maturity.

#### 4. SOME IMPORTANT PRINCIPLES THAT CROP MODELS SHOULD CONSIDER FOR HYPOTHESIZING GENETIC IMPROVEMENT

To be truly useful for genetic improvement, crop models must realistically simulate crop physiological responses to environmental factors and how genetic variation affects the processes of crop carbon, water and nutrient balance. These principles will be discussed in three ways: (i) do the models have adequate mechanism to simulate the traits of interest; (ii) do the models honour the C, N, water and energy balance; (iii) do the models consider pleiotropic effects, e.g. physiological linkages.

The first important question is whether the current GSPs and the mechanisms in the crop models are sufficient to correctly simulate a given process and its associated genetics. Does the GSP actually represent the genetic trait? For example, RUE is an approach used in many crop models, but RUE is a complex trait affected by many genes (QTLs) and there is no direct connection to leaf-level physiology that a molecular geneticist can measure. In other words, RUE-based models may be too simple for some purposes. That would imply that crop models should have a minimum level of physiology, such as leaf photosynthesis, respiration and organ growth, to enable one to interpret genetic effects. That level of detail may be needed for the modeller to be able to communicate successfully with geneticists. As Parent and Tardieu (2014) stated, the phenotype (in this case variation in RUE) must be the emergent outcome of how genes express themselves in response to environment. Thus, RUE is an emergent outcome, not a genetic trait. This is also true for other GSPs, although some of them more closely describe physiological processes than RUE. A goal should be to incorporate genetic effects based on known or hypothesized pathways for their influence on physiological processes. For a molecular marker effect to be incorporated into a crop model, one should know the mode of action of that QTL. What is the specific mode of action and environmental sensitivity of a given QTL at its more basic level of action? Saying that the QTL affects RUE or yield or harvest index (HI) or seed size is not sufficient, as they are the emergent outcomes. Following the logic of Hammer et al. (2016) and Chenu et al. (2018), these are examples of complex traits that should be dissected

Table 1. Definitions of GSPs for the CROPGRO model with default values for dry bean (cv. Porrillo Sintetico), soybean (cv. MG 5 Hutcheson) and peanut (cv. Georgia Green). td, thermal days. ptd, photothermal days.

GSP name	GSP definition	Dry bean		Soybean		Peanut	
		Porrillo Sintetico	MG 5 Hutcheson	Porrillo Sintetico	MG 5 Hutcheson	Porrillo Sintetico	MG 5 Hutcheson
CSDL	Critical short daylength below which reproductive development progresses rapidly with no daylength effect (h)	13.50	12.58	13.50	12.58	11.84	11.84
PPSEN	Slope of the relative response of development to photoperiod with time (1/h)	0.017	0.311	0.017	0.311	0.000	0.000
EM-FL	Time from emergence to first flower appearance (ptd <sup>+</sup> )	31.5	22.0	31.5	22.0	21.2	21.2
FL-SH	Time from first flower to first pod (ptd <sup>+</sup> )	3.0	8.0	3.0	8.0	9.2	9.2
FL-SD	Time from first flower to first seed (ptd <sup>+</sup> )	8.0	15.5	8.0	15.5	18.8	18.8
SD-PM	Time from first seed to physiological maturity (ptd <sup>+</sup> )	20.5	35.0	20.5	35.0	74.3	74.3
FL-VS	Time from first flower to last leaf on main axis – from Ecotype file (ptd <sup>+</sup> )	11.0	9.0	11.0	9.0	68.0	68.0
FL-LF	Time from first flower to end of leaf expansion (ptd <sup>+</sup> )	21.0	18.0	21.0	18.0	85.0	85.0
TRIFOL	Rate of node appearance at opt. temp (node per td) – in Ecotype file	0.35	0.32	0.35	0.32	0.35	0.35
LFMAX	Maximum leaf photosynthetic rate at 30 °C, 350 ppm CO <sub>2</sub> and high light (mg CO <sub>2</sub> m <sup>-2</sup> s <sup>-1</sup> )	1.00	1.05	1.00	1.05	1.20	1.20
SLAVR	Specific leaf area of cultivar under standard growth conditions (cm <sup>2</sup> g <sup>-1</sup> )	300	400	300	400	270	270
SIZLF	Maximum size of full leaf (compound leaf) (cm <sup>2</sup> )	150	230	150	230	18	18
XFRT	Maximum fraction of daily growth partitioned to seed + shell	1.00	1.00	1.00	1.00	0.87	0.87
WTPSD	Genetic potential weight per seed (g)	0.23	0.18	0.23	0.18	0.69	0.69
SFDUR	Seed-filling duration for pod cohort (ptd <sup>+</sup> )	21.0	23.0	21.0	23.0	38.0	38.0
SDPDV	Seeds per pod, standard growth conditions (# per pod)	5.20	2.05	5.20	2.05	1.65	1.65
PODUR	Duration of pod addition (ptd <sup>+</sup> )	11.5	10.0	11.5	10.0	29.0	29.0
THRSH	Threshing percentage, maximum % of seed to seed + shell	78.0	78.0	78.0	78.0	80.0	80.0
SDPRO	Potential seed protein (fraction)	0.235	0.400	0.235	0.400	0.270	0.270
SDLIP	Potential seed lipid (fraction)	0.030	0.200	0.030	0.200	0.510	0.510

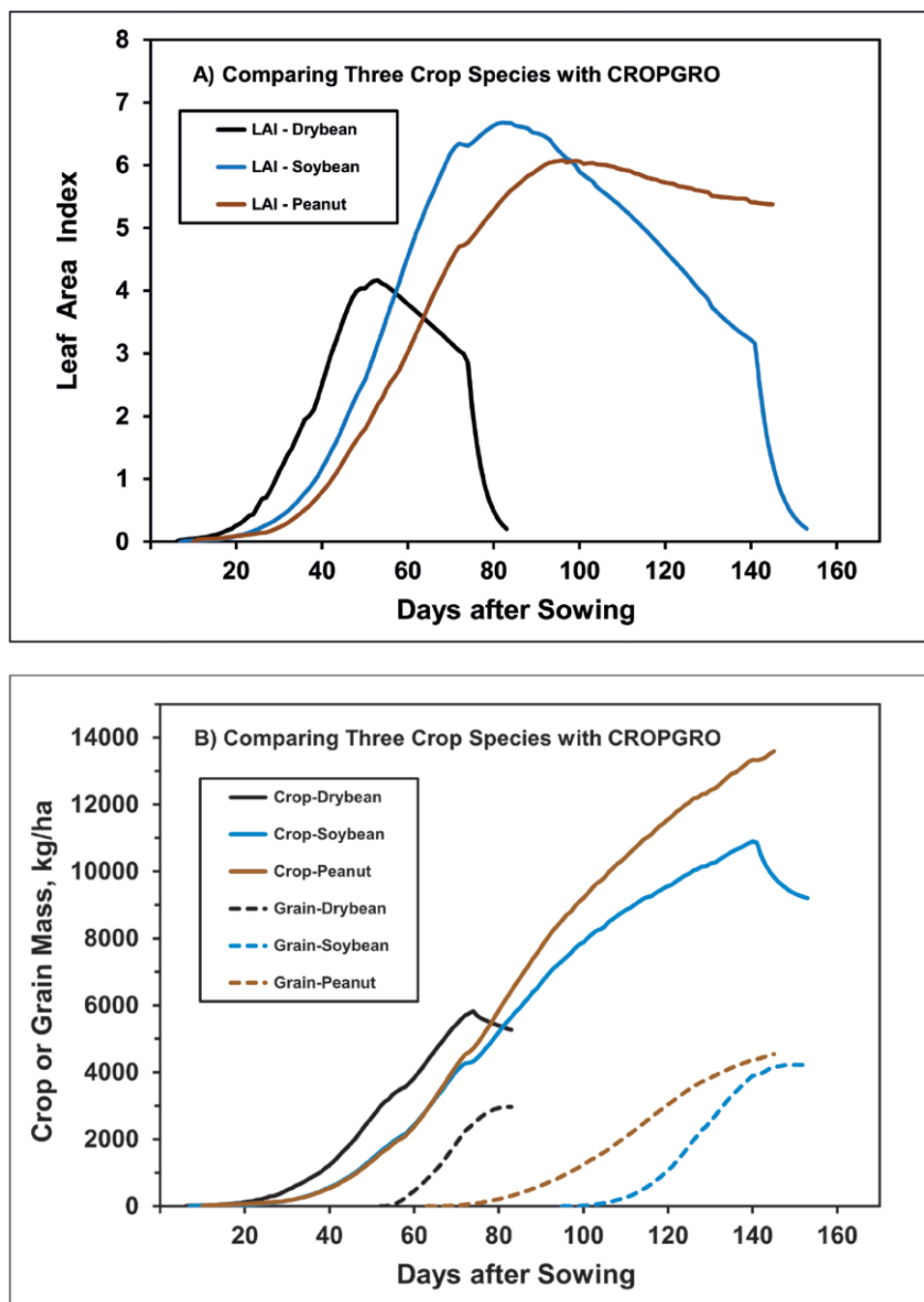


Figure 1. (A) Leaf area index, and (B) total crop mass and grain mass over time for dry bean cv. Porrillo Sintetico, soybean cv. Hutcheson and peanut cv. Georgia Green, sown 27 April 1999 at Plains, GA and grown under optimum irrigation.

into component traits at ecophysiology level which are then evaluated for their effects on those emergent outcomes.

A second aspect is whether the models honour the C balance, N balance, water balance and energy balance. Is there a free lunch? This is not an easy question to understand, because one might argue that increasing LAI is a simple way to increase early seedling vigour, but in fact, allocation to leaf area growth comes at the expense of less assimilate

allocation to roots, and there are also feed-forward effects to consider. Another example is that greater constitutive (all the time) allocation to roots may enhance drought tolerance but may come at the expense of shoot (and reproductive) growth. Water conservation associated with reduced transpiration (reduced leaf conductance) trait is proposed for drought conditions. But the 'no free lunch' problem to consider here is that reduced leaf conductance also reduces leaf photosynthesis, and

**Table 2.** Definitions of GSPs for the CERES-Maize and IXIM-Maize models with example values for three types of cultivars.

<sup>†</sup>Degree days above  $T_b = 8^\circ\text{C}$ . Compute with  $T_{\text{mean}} = (T_{\text{max}} + T_{\text{min}})/2$ . <sup>††</sup>Additional coefficients for IXIM-Maize. <sup>†††</sup>Simulated anthesis, maturity, grain yield and final biomass, for three hypothetical CERES-Maize cultivars sown 27 April 1999, at 7.5 plants per m<sup>2</sup>, 224 kg N ha<sup>-1</sup>, on a Clarion soil at Ames, IA.

GSP name	Cultivar coefficient definition	Short 2500 GDD	Medium 2600 GDD	Long 2700 GDD
P1	Thermal time from emergence to end of juvenile phase during which plant is not sensitive to photoperiod (GDD <sup>†</sup> )	160	185	212
P2	Days delay in development per hour increase in photoperiod above the critical short photoperiod (days/hour [d/h])	0.400	0.500	0.500
P5	Thermal time from silking to physiological maturity (GDD)	780	850	890
G2	Maximum possible number of grains (kernels) per plant	730	800	840
G3	Grain (kernel) filling rate during linear grain-filling under optimum conditions (mg per kernel per day)	8.6	8.5	8.5
PHINT	Phylochron interval, thermal time between successive leaf tip appearance (GDD)	38.9	38.9	38.9
AX <sup>††</sup>	Leaf area of largest leaf (cm <sup>2</sup> per leaf)	—	720	780
LX <sup>††</sup>	Leaf longevity of most long-lived leaf, from 50 % expansion to 50 % senescence (GDD)	—	800	800
Simulated output <sup>†††</sup>				
Time to anthesis (days after sowing)		70	73	78
Time to maturity (days after sowing)		121	129	142
Grain yield (kg ha <sup>-1</sup> )		10 702	12 746	14 048
Final total biomass (kg ha <sup>-1</sup> )		20 632	23 124	24 637

the degree of photosynthesis reduction is the critical issue to resolve. Messina *et al.* (2015) showed that a limited transpiration trait increased maize drought tolerance, resulting in a yield increase for dry environments, but a slight yield decrease for well-watered environments. Going one step further, the reduced conductance may warm the canopy above the optimum temperature for photosynthesis. N balance examples are also possible, and at the very least, models that lack a soil-plant N balance are not useful for hypothesizing genetic variation in environments where N is a major limiting factor.

A third consideration is whether the crop models consider built-in pleiotropic linkages. For example, increased leaf-level photosynthesis is commonly associated with increased specific leaf weight (SLW) or specific leaf nitrogen (SLN). That is what we would call a pleiotropic linkage because intrinsic rate per unit leaf mass or N mass is not increased. If modelled correctly, this may come at a cost. If the plant makes high SLW or SLN leaves, the amount of total leaf area will be lower and light capture will be less. The net benefit will depend on whether the increased leaf rate offsets the decreased light capture. An example will be given of this later.

## 5. THE ROLE OF ENVIRONMENT AND TRAIT RESPONSE TO ENVIRONMENT IS VERY IMPORTANT

The environment plays a major role in influencing ultimate plant phenotype (size, height, yield), and the phenotype is, or should be, an emergent outcome of how the individual genes and combinations of genes respond to environment. Hammer *et al.* (2010) indicate that

crop models should have algorithms for underlying processes that link to genetics and lead to simulated phenotypes as emergent properties. Likewise, Parent and Tardieu (2014) argued that phenotype should be an emergent outcome of the genetic variation in processes as affected by environment. An example of this is HI, which plant breeders may call a trait, but in reality HI is an emergent outcome of many genes that affect the timing, duration, grain-set, grain-growth rate and intensity of partitioning to reproductive growth (even those five items are not single gene actions). So, how can the crop models be enhanced to get at the controllers of the underlying processes, rather than thinking that emergent outcomes are the crop model 'traits'? For example, Raymond *et al.* (2003) and Chenu *et al.* (2009) simulated the effect of genetic markers linked to thermal rate of leaf area expansion and proposed individual QTLs for optimum leaf extension rate, for sensitivity to temperature, and sensitivity to vapour pressure deficit. It is important to highlight that before QTL/gene effects can be incorporated into crop models, the models need to be developed and sufficiently enhanced to allow scaling up the impact from the leaf level to the plant level (Chenu *et al.* 2008). In their case, this involved inclusion of an hourly meteorological module and a leaf area growth module that simulated rate of leaf initiation, leaf tip appearance, leaf ligule appearance and leaf expansion rate dependent on hourly temperature, hourly vapour pressure deficit, predawn soil water potential and leaf rank position. Other examples are the studies of gene effects on cultivar life cycle of dry bean (White and Hoogenboom 1996; Hoogenboom *et al.* 1997, 2004; Hoogenboom and White 2003) and soybean (Messina *et al.* 2006). For dry bean and soybean, these are well-researched genes that are known to respond differently to daylength

and temperature. Thus, the time to flower and the time to maturity, as well as final grain yield are emergent outcomes of individual genes and in combinations. A  $G \times E$  interaction is a probable outcome if a gene for strong daylength effect is present, but that will occur only if the crop is compared for short versus long daylength environments.

## 6.0 LINKING QTLS AND GENES TO GENOTYPE-SPECIFIC PARAMETERS IN CROPGRO

### 6.1 Linking QTL markers to GSPs in CROPGRO-Soybean

Messina et al. (2006) linked GSPs in the CROPGRO-Soybean model to genes with a multiple linear regression approach, where 1 or 0 represents the presence or absence of the dominant allele. They illustrated how six loci related to photoperiod sensitivity and determinacy (E1, E3, E3, E4, E5 and E7) in soybean could be translated into GSPs in the CROPGRO-Soybean model. For example, CSDL and emergence to flowering (EMFL) were defined as a function of number of dominant loci (NLOCI) and presence or absence (1 or 0) of critical loci, as follows:

$$\text{CSDL} = 14.33 - 0.44 \text{ NLOCI} + 0.27\text{E3} - 0.48\text{E5} + 0.18(\text{NLOCI})(\text{E5})$$

For the cultivar Savory this translates into the following value of 14.33

$$\begin{aligned} \text{CSDL (Savory)} &= 14.33 - 0.44(0) + 0.27(0) - 0.45(0) \\ &\quad + 0.18(1)(0) = 14.33 \end{aligned}$$

While for the cultivar Vinton 81 this translates into a value of 14.16

$$\begin{aligned} \text{CSDL (Vinton 81)} &= 14.33 - 0.44(1) + 0.27(1) - 0.45(0) \\ &\quad + 0.18(1)(0) = 14.16 \end{aligned}$$

The equation for ptd from EMFL is the following

$$\text{EMFL} = 20.77 + 2.1\text{E1} + 1.8\text{E3}$$

The QTL genetic markers associated with these genes were determined for a set of cultivars. In the example above, cultivar Savory is e1e2e3e4e5, and differs from Vinton 81 which is e1e2E3e4e5. Other cultivar coefficients, PPSEN, EM-FL, FL-SD, FL-VS, SD-PM, V1-JU and R1PRO, were also affected by some of the same loci and number of loci. Putting these all together, allowed a different phenotype to be simulated based on absence or presence of multiple dominant loci (such as E3). Then, with independent data on maturity and yield from seven trial locations in Illinois, USA, over 5 years, the CROPGRO-Soybean model with genetic coefficients based on E genes predicted 75 % of the variation in days to maturity and 54 % of variation in yield across the 35 site-year combinations (Messina et al. 2006).

### 6.2 Linking genetic architecture to GSPs in CROPGRO-Drybean to develop ideotypes

To connect genes to physiological processes, one can hypothesize ideotypes based on either dominant or recessive genes. White and Hoogenboom (1996) developed the first gene-based model GeneGro from the BEANGRO model (Hoogenboom et al. 1992, 1994). In

CSM-GeneGro six known genes were linked to the physiological processes of the BEANGRO model through the GSPs. These included the Ppd gene for basic photoperiod response, the Hr gene to enhance the effect of Ppd, the Fin gene for determinacy, the Fd gene for early flowering and maturity and three seed size genes, i.e. Ssz1, Ssz2 and Ssz3. The CSM-GeneGro-Drybean model was calibrated for 46 cultivars and 10 trials conducted in Canada, USA, Mexico, and Colombia. The same cultivars were then evaluated with an independent data set from 26 trials for the same locations for a total of 333 observations (Hoogenboom and White 2003; White and Hoogenboom 2003). Following calibration and evaluation, 96 genotypes were created for the CSM-GeneGro-Drybean model based on either dominant or recessive genes for Ppd \* Hr \* Fin \* Fd \* Ssz1 \* Ssz2 \* Ssz3. These 96 genotypes were then evaluated for dry bean production environments including Michigan, Idaho, and Washington, using standard management practices with at least 30 years of historical weather data to determine the  $G \times E$  interactions and to identify the best performing ideotypes for each location. The genotype with all genes dominant except for Ssz3 resulted in the highest mean yield. With respect to that ranking, Genotype 2 (111110), Genotype 1 (111111) and Genotype 9 (111011) ranked highest for Michigan, Genotype 9 (111011), Genotype 2 (111110) and Genotype 10 (111011) ranked highest for Twin Falls, ID, and Genotype 10 (111011), Genotype 9 (111011) and Genotype 2 (111110) ranged highest for Prosser, WA. For the three locations, two of the genotypes ranked among the top three. These environments differed in growing season temperature and, therefore, the Fd gene for early versus late flowering and the Ssz3 gene for seed size were the critical genes.

## 7. PREPARING TO SIMULATE GENETIC IMPROVEMENT WITH A CROP MODEL

### 7.1 Use a crop model suited to the task. Does the model have the desired genetic trait capability?

Modellers and geneticists need to understand the limitations of the crop model and determine whether the model, with its GSPs, has the ability to simulate desired genetic variations (QTL, genes) in a realistic way. For example, if the model cannot simulate heat stress, disease tolerance, salinity, water-logging, soil Al saturation or soil compaction, then conducting synthetic breeding studies of those issues is not possible. Furthermore, if the genetic traits are related to drought, N fertilization, temperature or daylength effects, then it is important that the models be tested beforehand and documented for their capabilities to simulate responses to those factors, especially with respect to actually simulating a soil water and nitrogen balance. If the target trait relates to variation in leaf-level photosynthesis, then the crop model should simulate leaf photosynthesis at the appropriate temporal scale and scaling up to canopy assimilation in a realistic way.

### 7.2 Clearly define the target environment (weather, soils and anticipated crop management, whether irrigated or rainfed, or well-fertilized or not).

Target environments will vary by region, rainfall, temperature, soils and management. It is important to carefully characterize the crop environment of interest, especially considering the major production areas the breeding is targeted for (Chenu 2015; Cooper et al. 2016). For example, the crop cycle may be defined by intermittent droughts,



or terminal drought, or limited by killing frosts, or by a desire for two (rice) crops in a season.

**7.3 Obtain data (phenological data, final yields, time-series data on biomass, LAI and reproductive growth) for baseline crop cultivars for the target regions along with the weather, soils and management inputs needed simulate the model. Do some initial calibration to start from a realistic yield level.**

**7.4 Consult with plant breeders and geneticists to jointly decide upon the genetic traits to vary (and likely range possible) to improve yield or modify life cycle or other goal.**

Chenu *et al.* (2018) proposed that consultation with breeders and geneticists should be iterative with feedback, to avoid misconceptions on what is possible. They recommended an integrated approach that combines insights from crop modelling, physiology, genetics and breeding to characterize traits contributing to yield gain for target environments. Furthermore, they recommended that complex traits such as RUE or HI or transpiration efficiency should be broken down into component traits that contribute to the complex traits, to allow improved physiological understanding.

It is important that the extent of variation in a genetic trait be limited to ranges documented in the literature or known by plant breeders/geneticists. Knowledge of the possible genetic variability will prevent simulating unrealistic results that are outside the range of genetic feasibility. For example, RUE is thought to be a relatively conservative trait. So, under that constraint, RUE should not be varied much, because models show large almost 1:1 increases in yield with increased RUE.

**7.5 Determine whether the GSPs and parameters available in the crop model have the ability to simulate those desired genetic traits in a realistic manner. This may require comparison to known response data.**

Are there linkages among traits? These may occur genetically or because of conservation of C and N. First, there are true genetic linkages (where two genes reside in close proximity on the chromosomes) and therefore the outcome illustrates true linkage. The *Fin* gene in common bean is an example of this. It affects determinacy as well as time to first flower appearance and is possibly thought to be just a single gene (Bhakta *et al.* 2017). But in a secondary viewpoint, there are pleiotropic linkages that can be understood from a physiological or C or N balance approach. For example, a 1 % increase in either seed protein or seed oil for soybean is simulated to reduce yield by about 0.6–0.8 %, simply because of the additional energy needed to produce proteins in contrast to producing starch compounds that require lower energy to produce (Boote and Tollenaar 1994). Seed protein concentrations of soybean cultivars have declined by about two absolute percentage units over the past three decades in accordance with cultivar release date even while grain yield has increased (Naeve and Miller-Garvin 2019, US Soy Quality Report). It is possible that plant breeders

selected for yield without paying sufficient attention to maintaining seed protein. Hammer *et al.* (2016) provide an example of pleiotropic linkage related to both a physiological trait and to the conservation balance of N, in which the stay-green outcome of sorghum (*Sorghum bicolor*) can be a simulated outcome of a dwarfing trait which results in less N in stem, but more N in leaf which sustains LAI longer during grain-filling. Likewise, pleiotropic linkage of a physiological trait and water balance/conservation can be demonstrated by simulated tolerance to terminal drought being an emergent outcome of lower early leaf area (less tillering in sorghum, Hammer *et al.* 2016) or less early plant vigour in wheat (Bourgault *et al.* 2020).

Another example of pleiotropic linkage not caused by ‘pure’ genetics, but by C balance and leaf physics, is the strong relationship of increased leaf photosynthesis ( $A_{\max}$ ) to increased SLW which occurs in many species including soybean (Dornhoff and Shibles 1970; Buttery *et al.* 1981; Morrison *et al.* 1999). This creates a pleiotropic linkage not caused by ‘specific’ genetics, but it affects seasonal LAI, light interception and canopy assimilation. As simulated by Boote *et al.* (2003), increasing leaf photosynthesis with strict coupling to SLW gives much less benefit to canopy photosynthesis and yield than a pure increase in leaf rate not coupled to SLW. The reason is that increases in SLW increase leaf photosynthesis rate but reduce leaf area expansion and early season LAI. The net effect on growth and yield is complex and interactive effects occur depending on row spacing, sowing density, crop life cycle and elevated CO<sub>2</sub> (Boote *et al.* 2003; Boote 2011).

The ‘limited transpiration’ traits proposed by Sinclair *et al.* (2010) and Gilbert *et al.* (2011) provide another example of whether models are able to realistically simulate a trait. In one case the limited transpiration is routinely proposed to occur relatively soon when the crop experiences mild stress, while in another case limited transpiration is induced only under high VPD. These traits can give complex responses, because they influence C balance, water balance and energy balance, and additionally depend on time-series timing of rainfall events and soil water depletion. At its core, the action of the constitutive ‘all-the-time’ limited transpiration trait is that a small reduction in leaf conductance will reduce leaf transpiration and canopy transpiration, thereby allowing some soil water conservation prior to water deficit periods. The water conservation effect is hypothesized to maintain or increase yields because the crop extends better into a future rainfall period. But, an important issue here is that reduction in stomatal conductance also reduces leaf photosynthesis and the extent of the reduction in leaf rate is the crux of the problem. The reduction of photosynthesis (case 1) during good rainfall season may reduce yield, even while the same trait may increase yield under severe drought seasons. This is what Battisti *et al.* (2017) found with that approach in CROPGRO-Soybean, although Sinclair *et al.* (2010) with a simpler model (SSM) proposed greater benefits under water-limitation and very small negative effect under good rainfall environment. Both of these simulated cases represent G × E interactions (*same gene action, but effect depends on environment*). Case 2 with the photosynthetic reduction under high VPD is a more adaptive response since it does not act under moderate evaporative demand environments. Nevertheless, the reduction in photosynthesis associated with reduced conductance (transpiration) under high VPD may

approach unrealistic simulations of internal  $\text{CO}_2$  ( $C_i$ ). The data of Gilbert et al. (2011), if taken to the stronger conductance sensitivities to VPD, can give an internal  $C_i/C_a$  ratio approaching 0.4 (like C-4 crops) under high VPD which is a concern because that has not been reported previously and tends to violate theory. As an independent example of this, Cuadra et al. (2021) incorporated the high VPD effect on photosynthesis and conductance into the hourly energy balance version of CROPGRO, and reductions in canopy transpiration were obtained, although the benefits to growth and yield have not yet been investigated. In that hourly energy balance model, with strong sensitivities of conductance to high VPD using the Ball–Barry–Leuning sensitivity to VPD (Leuning 1995; Miner et al. 2017), the simulated  $C_i/C_a$  ratio fell as low as 0.4 under extreme large VPD, although it was between 0.6 to 0.8 most of the time. The message is that the modellers and their models must honour water balance (nothing is free) and resulting effects on crop C balance and energy balance (canopy temperature rises under large VPD) which pushes canopy transpiration back up somewhat. Sinclair et al. (2010) conducted simulations illustrating that reduction in transpiration at high VPD gave a soybean yield gain near 200 kg ha<sup>-1</sup> in the USA. Battisti et al. (2017) using a similar approach with the CROPGRO-Soybean model simulated a gain of 1–75 kg ha<sup>-1</sup> for most regions in Brazil, although the gain was 75–150 kg ha<sup>-1</sup> for some regions. He suggested that differences between the two examples could be attributed to either climate of the regions or the level of model penalization on transpiration and photosynthesis.

### 7.6 Conduct the hypothetical trait simulations for multi-year weather across differing target environments (weather, soils and management) to determine the genetic improvement and any genotype by environment by management interactions. Examples follow.

## 8. IN SILICO MODELLING OF TRAITS FOR GENETIC IMPROVEMENT

Assume that you have defined the target environment (multiple years of weather for a site and the soils), along with the basic cultivar type to define season length and the crop management (rainfed or irrigated, and N fertilization level). A crop simulation model for the given crop would be simulated with those soils and 20–30 weather years, with a baseline cultivar and crop management. The outcome would be a mean yield with the distribution of yields depending on weather variability. This defines your baseline yield case. Ideally, one should have yield trial data that verify that the model simulations compare reasonably to field-observed yields for the baseline cultivar. It is important to know that the future hypothetical evaluations are well-grounded. Caution is needed if yield trial data experience pest and disease losses, because most crop models do not account for pest, disease and other biotic stress effects.

Next, one would create hypothetical genetic traits. Assume that one wants the same season length, with earlier anthesis but longer grain-filling duration. This would require decreasing the GSP input of photothermal unit requirement to anthesis, and increasing the trait for

photothermal units from anthesis to physiological maturity. These two traits would be considered as ‘genetic coefficients’ for the crop model. The model would be simulated for the new virtual cultivar with the same weather and soils, to see how the two newly modified traits affect the percent change in mean yield, as well as the resulting yield variability.

The simulations should be conducted with multiple target environments, to evaluate whether the new trait is best in just some environments or all environments (Löffler et al. 2005; Putto et al. 2008; Loison et al. 2017). This would be equivalent to testing for a G × E effect. The target environments should be described in terms of weather (over multiple seasons), water availability, soil physical and chemical constraints, desired crop life cycle and even management constraints. The target environments can be: (i) low N or high N level, (ii) rainfed versus irrigated, (iii) altered management, (iv) current versus future climate (higher  $\text{CO}_2$ , increased temperature or altered rainfall). The advantage is that virtual crop modelling can quickly evaluate trait effects on yield responses for multiple weather and soil environments (or future climate change), without actually doing the multi-location field trials. This can save years and millions of dollars in conducting multi-location trials. However, there are important things to consider when these computer experiment exercises are conducted: (i) What is the degree of confidence in the crop model ability to simulate that trait or to simulate yield response to weather and management generally? (ii) What is the knowledge of the range of genetic variability in the trait that one is simulating and what is possible?

### 8.1 Evaluating possible range of a GSP compared to literature-reported range

Crop models can be used to evaluate the quantitative response over the entire possible range of a single trait, such as rate of root depth progression. Crop models can also be used to evaluate response to multiple combinations of traits. In such evaluations, it is essential that the feasible genetic range for each given GSP be considered relative to reported literature. Figure 2 from Boote et al. (2003) illustrates soybean yield response to simulated rate of root depth progression for MG 3 soybean grown on a deep Nicollet clay loam at Ames, IA. The published range of variation in this trait comes from Kaspar et al. (1978, 1984). This figure illustrates several important points: (i) that the response to a given trait over its whole range may become asymptotic or saturating (responses are not necessarily linear), (ii) that the response within the feasible published range of variation may be relatively modest as shown here and (iii) that yield variability may be influenced by the trait, in this case, deeper rooting leads to less variability in yield.

The full range of yield response to another frequently studied genetic trait, light-saturated leaf photosynthesis (LFMAX) is illustrated in Fig. 3 from Boote et al. (2003). The published literature for soybean indicates that most of the genotypic increase in light-saturated leaf rate is associated with increase in SLW (or reciprocal of SLA) (Dornhoff and Shibbes 1970; Buttery et al. 1981; Morrison et al. 1999). So, one cannot simply propose increases in leaf photosynthesis without considering this pleiotropic relationship to SLW. In other words, there is a cost (of crop dry matter) to pay for making thicker leaves, which results in smaller canopy LAI and reduced light interception. As a result, the optimistic picture of increasing yield with increasing

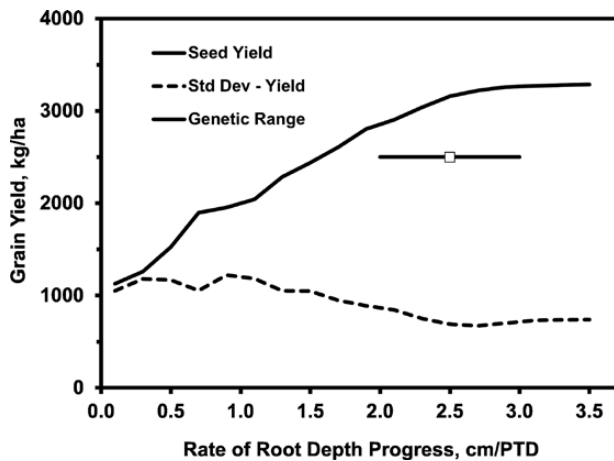


Figure 2. Simulated soybean yield (solid line) and standard deviation (dashed line) as affected by average simulated rate of root depth progression over 17 rainfed seasons at Ames, IA. Horizontal dashed bar represents published genetic range for rate of root depth progression. From [Boote et al. \(2003\)](#).

single leaf photosynthesis is made more complex by this pleiotropic connection of leaf photosynthesis to SLW. In [Fig. 3](#), the yield response to the LFMAX with coupling to SLW is really quite modest (rising, but only slowly), while the response to LFMAX with no coupling to SLW is much more optimistic and is only slowly asymptotic. Also important is the range of measured genetic variation in light-saturated LFMAX of soybean (shown as a horizontal bar in [Fig. 3](#)) shows relatively modest potential for improving yield based on increased leaf photosynthesis. The net effect (of increased SLW to obtain increased LFMAX) on growth and yield is complex and interactive effects occur depending on row spacing, sowing density, crop life cycle and elevated  $\text{CO}_2$  ([Boote et al. 2003](#); [Boote 2011](#)). The increased SLW trait can give neutral or negative effects under low sowing density, wide row spacing and ambient  $\text{CO}_2$ , but becomes increasingly more positive for yield at narrow row spacing, high sowing density and elevated  $\text{CO}_2$  (all of which compensate for the reduced LAI).

## 8.2 Evaluating single and multiple combinations of traits for genetic yield improvement

We can use crop models to evaluate single effects of traits as well as combinations of multiple genetic traits to improve yield. Plant breeders, in their selection process, will happily accept all traits that contribute to increased yield unless quality and disease resistance are compromised. Particularly with past traditional plant breeding, breeders may not have known all the specific traits of their improved cultivars, or what the traits may be contributing, but they were happy to accept the improvement in yield. Thus, we need to consider that crop modelling for yield improvement should also consider the degree of additivity of multiple traits. It is important to evaluate the traits in different environments, e.g. water-limited, elevated temperature, elevated  $\text{CO}_2$  and management conditions, to appreciate  $G \times E$  interactions, as shown in the simulation examples below.

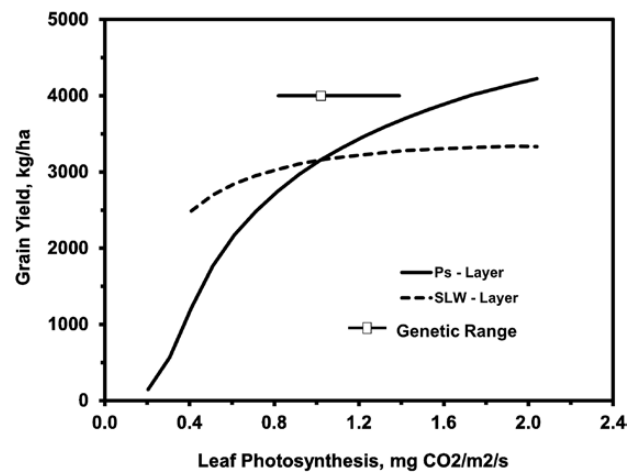


Figure 3. Simulated soybean yield response to LFMAX, coupled (dashed line), and not coupled (solid line), to SLW. Averaged over 17 rainfed seasons at Ames, IA. Horizontal bar represents published genetic range. From [Boote et al. \(2003\)](#).

[Boote et al. \(2001\)](#) analysed the growth patterns of old and new soybean cultivars and concluded that there were multiple traits (up to five or six, as interpreted from the CROPGRO-Soybean analysis), which contributed to the observed yield gains of 12–23 % for the improved cultivars. These traits included earlier onset of pod-set, faster pod addition, longer grain-filling, increased leaf photosynthesis and slower N mobilization. Later crop model simulations by [Boote \(2011\)](#) evaluated the effect of combining multiple traits on soybean yield and suggested that the effects of traits were generally additive ([Table 3](#)). Of particular interest is that some traits were management- and climate-dependent. For example, some traits such as determinacy, early anthesis and increased SLW gave modest or no yield improvements in ambient  $\text{CO}_2$  and low management (wide rows, low sowing density, low input) conditions, but the same traits when placed in high  $\text{CO}_2$  and/or high management were contributors to increased yield. Traits such as early anthesis and determinacy that reduce LAI are not a disadvantage under conditions that favour abundant vegetative growth. This message is particularly important for breeding for crop yield improvement under present and future  $\text{CO}_2$  increase. In fact, some GSPs gave a greater response under elevated  $\text{CO}_2$  than under 350 ppm  $\text{CO}_2$  (see highlighted increases in **bold** in [Table 3](#)). The additivity of multiple GSPs in two-, three- and four-way combinations progressively increased yield above the base MG 3 cultivar. A single trait of longer filling period gave 5.3 % yield increase, filling period plus high SLW gave 6.3 % and filling period plus slower N mobilization gave 10.2 % increase in yield (these are conservative possibilities). Staying within feasible genetic range, the three-way combination of 15 % longer filling period, 10 % higher LFMAX and 10 % slower N mobilization from vegetative organs gave 17.7 % increase in yield. Pushing the envelope for somewhat greater changes (such as 15 % increase in LFMAX) in three- or four-way combinations resulted in a yield increase of 20.6–23.9 %, approaching what the plant breeders have achieved since the Williams 82 cultivar was released in

**Table 3.** Simulated soybean yield with GSP trait modifications with CO<sub>2</sub> rise from 350 to 500 ppm at Ames, IA with long-term historical weather (20 years) using CSM-CROPGRO-Soybean. Sown 1 May at 25 plants per m<sup>2</sup> with a 0.20-m row spacing on a Nicollet clay loam with 2.0-m deep rooting profile. Adapted from Boote (2011). Values in bold illustrate genotype by environment interactions for a given trait or combination of traits.

Genetic trait	Carbon dioxide concentration			
	350 ppm		500 ppm	
	kg ha <sup>-1</sup>	% increase	kg ha <sup>-1</sup>	% increase
Standard (MG 3)	3403		4084	
Determinate (Dt)	3413	0.30		<b>2.25</b>
+10 % filling period (FP)	3583	5.29		<b>6.00</b>
+10 % SLW	3439	1.06		1.37
+10 % LFMAX	3605	5.94		4.93
10 % slower N mobilization	3517	3.36		2.81
Dt + 10 % FP	3481	2.30		<b>5.26</b>
Dt + 10 % FP + 10 % SLW	3496	2.74		<b>5.81</b>
Dt + 10 % FP + 10 % LFMAX	3709	9.00		<b>11.30</b>
10 % FP + 10 % SLW	3619	6.34		7.31
10 % FP + 10 % LFMAX	3800	11.67		11.43
Dt + 10 % FP + Nmob	3629	6.65		<b>9.71</b>
Dt + 10 % FP + 10 % SLW + Nmob	3650	7.26		<b>10.27</b>
Dt + 10 % FP + 10 % LFMAX + Nmob	3869	13.70		<b>15.82</b>
10 % FP + 10 % SLW + Nmob	3750	10.19		10.87
10 % FP + 10 % LFMAX + Nmob	3920	15.19		14.77
+15 % FP + 10 % LFMAX + Nmob	4006	17.73		17.45
+15 % FP + 15 % LFMAX + Nmob	4104	20.59		19.84
+15 % FP + 15 % LFMAX + Nmob and +10 % leaf N conc	4217	23.93		21.82

1982 (Williams 82 is the baseline in Table 3, which also corresponds approximately to the period when CO<sub>2</sub> was about 350 ppm). The responses to GSPs and the additivity of GSPs continued to be present at the higher CO<sub>2</sub> concentration of 500 ppm.

Yield improvements with single and multiple GSP trait variations were simulated with the CROPGRO-Peanut model using 15 years of rainfed conditions in Ghana under hypothesized increase in CO<sub>2</sub> and a +3 °C increase in temperature (Table 4). The short-season cultivar, Chinese, had been well-calibrated for multiple seasons in Ghana and the soil water-holding characteristics were also well-established by comparison to observed data (Naab et al. 2004; Naab et al. 2005). Because this is a short-season cultivar with a relatively low LAI, yield increased 4.2 % with delayed flowering (+10 % EM-FL) and increased 9.0 % with longer grain-filling duration (+10 % SD-PM), and 12.4 % with both traits. Likewise, with a relatively low LAI, an increase in photosynthesis (+10 % LFMAX) caused a 7.3 % increase in yield, while an increase in SLA (decreased SLW) increased yield by only 0.5 %. An increase in partitioning intensity increased yield by 3.4 % because this cultivar is relatively indeterminate. Moreover, combinations of traits such as delayed flowering, longer grain-fill, higher partitioning (+10 % XFRT) and higher photosynthesis were additive and increased yield by 24.7 %. Of particular interest is that three traits (delayed flowering, longer grain-filling and increased partitioning) increased yield more under the +3 °C temperature scenario than at ambient conditions. This would be a G × E interaction. Under future

climate change, a temperature increase of this magnitude is probable and would cause a shorter life cycle, which the increased EM-FL and increased SD-PM help to offset. Zheng et al. (2016) also suggested that genetic alleles for longer life cycle would be needed for crops such as wheat under the rising temperature anticipated with climate change. In addition, the peanut model simulates a reduced rate of pod addition and partitioning under elevated temperature (relationship in model is based on observations under elevated temperature (Prasad et al. 2003; Boote et al. 2018)), and the increased partitioning helps to bring that back.

Crop growth models can be used to investigate traits contributing to past genetic yield improvement by comparison to phenotyping data collected on old versus improved cultivars. This was accomplished by Narh et al. (2015) who evaluated improved peanut cultivars released by ICRISAT for use in West Africa, by comparison to cv. Chinese the baseline cultivar in that region. Narh et al. (2015) used the CSM-CROPGRO-Peanut model as an analysis tool to evaluate possible genetic contributions to yield increase among a set of 19 cultivars that included improved cultivars from ICRISAT plus local cultivars. Data on phenology, biomass, pod mass, pod HI and final yield were collected on 19 cultivars over two seasons at four sites in Burkina Faso and Ghana. Optimization techniques were used with the model to solve GSPs. The two highest yielding lines Nkatesari and ICGV-IS 96814, both ICRISAT-released genotypes, yielded 76–80 % more than Chinese. The model was able



to successfully mimic this yield increase with the GSPs described in Table 5. Traits important to yield under this no-fungicide rainfed production included longer life cycle, higher photosynthesis, higher partitioning, longer grain-fill and modest leafspot resistance. Leafspot resistance was partial, but certainly contributed to the higher leaf photosynthesis (greater crop growth rate) along with extended life cycle. The longer life cycle, longer grain-filling and higher partitioning were also documented in the phenology and pod HI observations.

### 8.3 Simulating genotype by environment interactions

Genotype by environment interactions are of great interest to plant breeders because this helps select cultivars and cultivar traits for different target environments. Table 6 illustrates strong  $G \times E$

interactions for the case of simulated GSP traits with the CSM-CROPGRO-Chickpea model for crops grown either under full irrigation or under water-limited terminal drought (Boote *et al.* 2013). Chickpea (*Cicer arietinum*) in India is typically sown at the end of the monsoon and depends on stored residual water on high-clay soils (Singh *et al.* 1999a, b). The target environment, irrigated versus rainfed, is important, because the responses to GSPs were often opposite and large for contrasting soil water availability. For example, the effect of increased SLW was to decrease yield 11 % under irrigation but to increase yield 18 % under water-limitation. Increased SLW had a negative effect on yield under irrigation because it reduced LAI and light interception. But the same higher SLW trait was beneficial under rainfed conditions because it reduced LAI, light interception and transpiration, thus conserving water for grain growth later in the life cycle. Later flowering acted

**Table 4. Simulated peanut seed yield response to GSP trait combinations under two CO<sub>2</sub> and temperature levels, grown rainfed with 15 years of weather at Wa, Ghana using CSM-CROPGRO-Peanut. Values in bold illustrate genotype by environment interactions for a given trait or combination of traits. <sup>†</sup>For definitions of GSP traits, see Table 1.**

Cultivar trait modified <sup>†</sup>	Ambient CO <sub>2</sub> 380 ppm	CO <sub>2</sub> 500 ppm	+3 °C temp	+3 °C 500 ppm –20 % RF
	% change			
Chinese (90-d std) (kg ha <sup>-1</sup> )	1841	2248	1553	1770
+10 % LFMAX (leaf Ps)	7.3	6.4	6.9	5.4
+10 % specific leaf area	0.5	0.5	0.5	0.0
+10 % EMFL	4.2	3.6	4.3	3.9
+10 % SDPM	9.0	8.3	<b>12.6</b>	<b>12.4</b>
+10 % EMFL & SDPM	12.4	11.1	<b>16.1</b>	<b>14.6</b>
+10 % WTPSD (wt/sd)	0.7	0.8	0.7	0.5
–10 % pod adding duration	0.1	0.7	3.9	3.6
+10 % XFRT (partitioning)	3.4	3.8	5.3	5.7
+10 % SIZLF (veg. vigour)	1.5	1.4	0.9	0.8
EMFL, SDPM, XFRT	16.5	15.5	<b>21.3</b>	<b>19.7</b>
EMFL, SDPM, LFMAX	20.4	18.0	22.7	20.3
EMFL, SDPM, XFRT, LFMAX	24.7	22.6	<b>28.6</b>	<b>25.8</b>

**Table 5. Observed pod yield and observed pod HI over two seasons at four sites, and GSPs of peanut cultivars derived from data using the CROPGRO-Peanut model with an optimizer (from Narh *et al.* 2014, 2015). Only the three highest yielding cultivars and the three local check cultivars are shown here. <sup>†</sup>PD, photothermal days.**

Cultivar	Observed pod yield (kg ha <sup>-1</sup> )	Observed pod HI	Cultivar GSP traits <sup>†</sup>							
			LFMAX	XFRT (fraction)	PODUR (PD)	SFDUR (PD)	THRSH (%)	EM-FL (PD)	FL-SD (PD)	SD-PM (PD)
ICGV-IS 96814	1760a	0.405a	1.60	0.72	20.7	25.0	72.8	21.0	23.4	50.70
NKATESARI	1720a	0.400ab	1.57	0.70	15.0	23.0	73.9	20.0	25.0	50.00
ICGV-IS 92093	1580b	0.425a	1.55	0.71	18.0	24.0	74.9	21.0	26.0	48.00
CHINESE	980def	0.330de	1.17	0.59	7.5	20.0	71.5	19.9	22.7	35.14
DOUMBALA	960d-g	0.355cd	1.16	0.58	9.5	22.5	70.0	19.0	20.0	39.50
TS 32-1	890fh	0.370bc	1.10	0.59	8.0	24.0	72.0	20.0	20.5	38.20

**Table 6.** Grain yield response to varying GSPs, simulated for 22 years for Annigeri chickpea grown under either rainfed or irrigated conditions at Patancheru, India. Sown on 29 September (day 302) on a very fine montmorillonitic clay soil, starting at field capacity. Simulated with the CSM-CROPGRO-Chickpea model as developed by Singh and Virmani (1994) and modified by Singh et al. (2014). Table used with permission from Boote et al. (2013).

Cultivar GSP modified	Rainfed		Irrigated	
	Mean yield kg ha <sup>-1</sup>	Percent change %	Mean yield kg ha <sup>-1</sup>	Percent change %
Standard simulation (Annigeri)	773		2614	
Rooting/SLW/LFMAX traits				
+10 %, rate of root depth progression	791	2.3	2614	0.0
+10 % leaf photosynthesis (LFMAX)	783	1.3	2961	13.3
+10 % specific leaf weight (SLW)	916	<b>18.5</b>	2328	<b>-10.9</b>
Life cycle traits				
10 % longer from emergence to anthesis	669	<b>-13.5</b>	3016	<b>15.4</b>
10 % longer seed-fill (first seed to maturity)	787	1.8	2893	10.7
Seed size/partitioning				
10 % larger potential seed size	746	<b>-3.5</b>	2709	<b>3.6</b>
10 % faster pod addition	774	0.2	2717	3.9

to increase LAI, and gave 15.4 % yield increase under irrigation, but reduced yield 13.5 % under the terminal drought. The common factor was the amount of LAI produced and amount of soil water conserved and left at the time seed growth began. Likewise, higher photosynthesis had a much larger benefit under irrigation; while under terminal drought, it had little benefit because it increased early LAI and water extraction too much. Bourgault et al. (2020) suggested that an early vigour trait in wheat could similarly predispose the crop for yield reduction under terminal drought because of early depletion of soil water caused by the earlier LAI.

It is important to understand how and when crop models reproduce  $G \times E$  interactions. It is proposed that  $G \times E$  interaction can come from a 'normal' single gene or gene package where the effect of that gene is neutral in one environment but negative or positive in another environment, as in the example for chickpea grown under irrigation or terminal water deficit. The point here is that there is not a need for a 'special'  $G \times E$  'gene', but rather that a given gene or gene package may be beneficial in one environment, but negative in another, as also suggested by Chapman (2008), Chenu et al. (2009), Tardieu (2012) and Hammner et al. (2016).

In the examples so far, we have illustrated that simulated  $G \times E$  can be an emergent outcome for environments differing in water availability, CO<sub>2</sub>, temperature and crop management.  $G \times E$  effect for locations (L) can occur if the location varies in soil water-holding capacity, rainfall and temperature. Putto et al. (2013) found  $G \times L$  effects in different locations as GSPs were varied with the CROPGRO-Peanut model.  $G \times E$  interaction for management can occur if the management varies in irrigation, N fertilization, row spacing and plant population. For all the water supply-based environments, deep rooting trait will benefit only if water deficit occurs. Even different cultivar life cycles can result in a  $G \times E$  interaction if the environment is different during the non-common part of the life cycle.

#### 8.4 Simulating adaptation to climate change factors

Crop models are useful for evaluating genetic traits for adaptation to climate change factors (Boote 2011; Singh et al. 2012, 2017; Hammer et al. 2020). In a project funded by the International Food Policy Research Institute, K.J.B. modified the code of the CERES-Pearl Millet model to account for elevated temperature effects on grain-set. Then, the model was calibrated by Singh et al. (2017) to pearl millet (*Pennisetum glaucum*) data from six sites in India and two sites in West Africa, followed by simulating response to climate change scenarios (increased temperature, rainfall change, CO<sub>2</sub> increase) for those sites (Table 7). Traits involved a 10 % shorter life cycle, a 10 % longer life cycle, increased productivity, deeper more effective water extraction and heat tolerance for grain-set and single grain-growth rate. Life cycle was modified by changing P1, P2O and P5 (definitions are similar to those of CERES-Maize in Table 2). Simulated productivity trait was accomplished by 10 % increases in G1, G4, GT and RUE parameters. For drought tolerance, the shape of rooting depth was made deeper and the lower limit of the soil water-holding traits was reduced to give 5 % more available soil water. The grain-set sensitivity to temperature was based on field observations of grain-set at an elevated temperature location (Gupta et al. 2015). For the model, the temperature thresholds were set to give no reduction in grain-set below 33 °C, but with a linear reduction in grain number between 33 to 39 °C, and zero grain-set at 39 °C. The grain number in the millet model is determined by the assimilate supply and cumulative biomass, and the daily mean temperature during the ISTAGE4 phase. To hypothesize a heat tolerance trait for an improved cultivar, both the threshold and the ceiling failure temperature for grain-set were increased by 2 °C to 35 and 41 °C.

Simulations with the CERES-Pearl Millet model for the sites in India (Hisar, Jaipur, Jodhpur, Bikaner, Aurangabad and Bijapur) and in West Africa (Sadore and Cinzana) revealed that a shorter life cycle was not a successful strategy under climate change, although a longer life cycle increased yield for about half of the sites (Table 7).

**Table 7.** Yield of base cultivar under climate change (CG) by 2050 and percentage gain or loss in yield for virtual pearl millet cultivars with 10 % shorter duration, 10 % longer duration, yield potential (increased productivity), drought tolerance (DT), heat tolerance (HT) or drought plus heat tolerance (DT + HT) traits at sites in India and West Africa. Yield changes of 6 % or less were deemed not significant.

Location	Season (CG) temperature $T_{\max}/T_{\min}$	Base cultivar yield	Shorter duration	Longer duration	Yield potential	Drought tolerant	Heat tolerant	DT + HT
	°C	kg ha <sup>-1</sup>	Percent gain or loss in yield					
Hisar	39.0/25.0	1105	-18	17	25-30	5-9	7-8	13-17
Jaipur	38.0/26.0	1303	-33	4	13-43	3-8	6-7	9-15
Jodhpur	38.5/27.2	1531	-16	6	18-24	4-13	8	13-22
Bikaner	40.3/27.6	544	-7	0	19-20	15-22	16-18	36-42
Aurangabad	33.6/24.0	1925	-42	46	31-43	5-9	0-1	6-10
Bijapur	33.4/24.6	1729	-47	46	33-52	7-14	1	9-15
Sadore	37.5/28.7	1018	-17	-6	12-21	6-25	7-17	16-46
Cinzana	36.5/27.2	1375	-29	4	10-21	8-9	6-7	14-17

The increased productivity trait (four combined GSPs listed above) increased yield at most sites, but were not uniformly consistent across all sites, but we have concerns for the reality of modifying all those traits at once, and whether plant breeders can achieve them. The drought tolerance trait (deeper more effective water extraction) gave more response for the low-yielding water-limited sites of Bikaner and Sadore. The heat tolerance trait (+2 °C more tolerance) gave no yield benefit for the cool sites (Aurangabad, and Bijapur), but gave 17–18 % increases for the warmest sites. This would be a G × E interaction. The simulated combination of droughts and heat tolerance traits was additive. In a similar manner, Singh *et al.* (2012) evaluated GSPs of CROPGRO-Peanut for peanut adaptation to climate change across multiple sites in India. A 2 °C increase in heat tolerance of pod-set and seed growth rate resulted in larger yield responses (3–12 %) under climate change than under current climate, except for two cool sites where there was little difference. Responses to multiple traits were also generally additive.

## 9. LOOKING TO THE FUTURE ROLE OF CROP MODELS IN GENETIC IMPROVEMENT

There is great opportunity to use dynamic crop simulation models as tools to evaluate past genetic improvement, to evaluate virtual cultivars for future yield improvement in target environments, to hypothesize traits that account for G × E responses in many environments and to assist plant breeder trait selection for yield improvement. We agree whole-heartedly with Muller and Martre (2019) that crop models are important tools at the cross-roads needed to link physiology, genetics and phenomics. With the rapid advance in genotyping and QTL analysis, the most limiting factor now is the phenotyping of crop traits and performance in multiple environments (Furbank and Tester 2011). The significant amount of resources directed to phenotyping and QTL analysis (Tardieu *et al.* 2017) will benefit considerably if combined with crop simulation modelling and physiological understanding as a way to integrate those many phenotypic outcomes over multiple environments (Chenu *et al.* 2018). There is good opportunity to integrate crop growth models with genome wide prediction to improve

genomic prediction accuracy, in particular when G × E interactions are an important determinant of performance (Technow *et al.* 2015; Cooper *et al.* 2016; Messina *et al.* 2018).

However, for linking crop models to genes and QTLs, it is important to consider the correct physiological mode of action and to consult with plant breeders on the feasible genetic range for a given trait. Most traits of interest to plant breeders are emergent outcomes of multiple genes and physiological processes, which implies that crop models may need improvement to include more detailed representation of processes and dissection of traits into component traits at ecophysiology level to better simulate those emergent outcomes (Hammer *et al.* 2016; Chenu *et al.* 2018). Evaluating traits with crop growth models reveals that G × E interactions are associated with environments differing in water supply, temperature, CO<sub>2</sub> level, soil water-holding characteristics and crop management. Models often mimic G × E because a single trait (gene) that is beneficial in one environment may be negative in another. The crop models will need a certain depth of physiological detail, genotypic information and understanding of genetic direct (G) and interactive (G × E and G × G) effects on dynamic physiological processes to robustly incorporate QTLs and genes into dynamic crop models.

## LITERATURE CITED

- Allen RG, Pereira LS, Raes D, Smith M. 1998. *Crop evapotranspiration. Guidelines for computing crop water requirements*. FAO irrigation and drainage paper n° 56. Rome, Italy: FAO.
- Battisti R, Sentelhas PC, Boote KJ, de S. Camara GM, Fraias JRB, Basso CJ. 2017. Assessment of soybean yield with altered water-related genetic improvement traits under climate change in Southern Brazil. *European Journal of Agronomy* 83:1–14.
- Bhakta MS, Gezan SA, Clavijo-Michelangelis JA, Carvalho M, Zhang L, Jones JW, Boote KJ, Correll MJ, Beaver J, Osorno JM, Colbert R, Rao I, Beebe S, Gonzalez A, Ricaurte J, Vallejos CE. 2017. A predictive model for time-to-flower in the common bean based on QTL and environmental variables. *G3: Genes, Genomes, Genetics* 7:3901–3912.
- Boote KJ. 2011. Improving soybean cultivars for adaptation to climate change and climate variability. Chapter 17. In: Yadav SS, Redden RJ,

- Hatfield JL, Lotze-Campen H, Hall AE, eds. *Crop adaptation to climate change*. West Sussex, UK: Wiley-Blackwell, 370–395.
- Boote KJ. 2019. The future of crop modelling for sustainable agriculture. In: Boote KJ, ed. *Advances in crop modelling for a sustainable agriculture*. Cambridge, UK: Burleigh Dodds Science Publishing, 477–507.
- Boote KJ, Pickering NB. 1994. Modeling photosynthesis of row crop canopies. *Hortscience* **29**:1423–1434.
- Boote KJ, Hoogenboom G, Jones JW, Ingram KT. 2009. Modeling N-fixation and its relationship to N uptake in the CROPGRO model. In: Ma L, Ahuja L, Bruulsema T, eds. *Quantifying and understanding plant nitrogen uptake for systems modeling*. Boca Raton, FL: Taylor & Francis Group LLC, 13–46.
- Boote KJ, Jones JW, Batchelor WD, Nafziger ED, Myers O. 2003. Genetic coefficients in the CROPGRO-soybean model: links to field performance and genomics. *Agronomy Journal* **95**:32–51.
- Boote KJ, Jones JW, Hoogenboom G, Pickering NB. 1998. The CROPGRO model for grain legumes. In: Tsuji GY, Hoogenboom G, Thornton PK, eds. *Understanding options for agricultural production*. Dordrecht, The Netherlands: Kluwer Academic Publishers, 99–128.
- Boote KJ, Jones JW, White JW, Asseng S, Lizaso JJ. 2013. Putting mechanisms into crop production models. *Plant, Cell & Environment* **36**:1658–1672.
- Boote KJ, Kropff MJ, Bindraban PS. 2001. Physiology and modeling of traits in crop plants: implications for genetic improvement. *Agricultural Systems* **70**:395–420.
- Boote KJ, Sau F, Hoogenboom G, Jones JW. 2008. Experience with water balance, evapotranspiration, and prediction of water stress effects in the CROPGRO model. In: Ahuja LR, Reddy VR, Saseendran SA, Yu Q, eds. *Response of crops to limited water: modeling water stress effects on plant growth processes, volume 1 of advances in agricultural systems modeling*. Madison, WI: ASA-CSSA-SSSA, 59–103.
- Boote KJ, Tollenaar M. 1994. Modeling genetic yield potential. In: Boote KJ, Bennett JM, Sinclair TR, Paulsen GM, eds. *Physiology and determination of crop yield*. Madison, WI: ASA-CSSA-SSSA, 533–565.
- Boote KJ, Vara Prasad Allen LH Jr, Singh P, Jones JW. 2018. Modeling sensitivity of grain yield to elevated temperature in the DSSAT crop models for peanut, soybean, dry bean, chickpea, sorghum, and millet. *European Journal of Agronomy* **100**:99–109.
- Bouman BAM, van Keulen H, van Laar HH, Rabbinge R. 1996. The 'School of de Wit' crop growth simulation models: a pedigree and historical overview. *Agricultural Systems* **52**:171–198.
- Bourgault M, Webber HA, Chenu K, O'Leary GJ, Gaiser T, Siebert S, Dreccer F, Huth N, Fitzgerald GJ, Tausz M, Ewert F. 2020. Early vigour in wheat: could it lead to more severe terminal drought stress under elevated atmospheric [CO<sub>2</sub>] and semi-arid conditions? *Global Change Biology* **26**:4079–4093.
- Buttery BR, Buzzell RJ, Findlay WL. 1981. Relationships among photosynthetic rate, bean yield and other characters in field-grown cultivars of soybean. *Canadian Journal of Plant Science* **61**:191–198.
- Chandler RF Jr. 1969. Plant morphology and stand geometry in relation to nitrogen. Chapter 12. In: Eastin JD, Haskins FA, Sullivan CY, Van Bavel CHM, eds. *Physiological aspects of crop yield*. Madison, WI: ASA-CSSA-SSSA, 265–285.
- Chapman SC. 2008. Use of crop models to understand genotype by environment interactions for drought in real-world and simulated plant breeding trials. *Euphytica* **161**:195–208.
- Chapman SC, Cooper M, Podlich D, Hammer GL. 2003. Evaluating plant breeding strategies by simulating gene action and dryland environment effects. *Agronomy Journal* **95**:99–113.
- Chenu K. 2015. Characterising the crop environment – nature, significance and applications. In: Sadras VO, Calderini DF, eds. *Crop physiology. Applications for genetic improvement and agronomy*, 2nd edn. Academic Press, 321–348.
- Chenu K, Chapman SC, Hammer GL, McLean G, Ben-Haj-Salah H, Tardieu F. 2008. Short-term responses of leaf growth rate to water deficit scale up to whole-plant and crop levels: an integrated modelling approach in maize. *Plant Cell and Environment* **31**:378–391.
- Chenu K, Chapman SC, Tardieu F, McLean G, Welcker C, Hammer GL. 2009. Simulating the yield impacts of organ-level quantitative trait loci associated with drought response in maize: a "gene-to-phenotype" modeling approach. *Genetics* **183**:1507–1523.
- Chenu K, Van Oosterom EJ, McLean G, Deifel KS, Fletcher A, Geetika G, Tirfessa A, Mace ES, Jordan DR, Sulman R, Hammer GL. 2018. Integrating modelling and phenotyping approaches to identify and screen complex traits: transpiration efficiency in cereals. *Journal of Experimental Botany* **69**:3181–3194.
- Cooper M, Podlich DW, Smith OS. 2005. Gene-to-phenotype models and complex trait genetics. *Australian Journal of Agricultural Research* **56**:895–918.
- Cooper M, Technow F, Messina C, Gho C, Totir LR. 2016. Use of crop growth models with whole-genome prediction: application to a maize multi-environment trial. *Crop Science* **56**:2141–2156.
- Cuadra SV, Kimball BA, Boote KJ, Suyker AE, Pickering N. 2021. Energy balance in the DSSAT-CSM-CROPGRO model. *Agricultural and Forest Meteorology* **297**. doi:10.1016/j.agrformet.2020.108241.
- Curry RB, Baker CH, Streeter JG. 1975. SOYMOD I: a dynamic simulator of soybean growth and development. *Transactions of the ASAE* **18**:963–968.
- Donald CM. 1968. The breeding of crop ideotypes. *Euphytica* **17**:193–211.
- Dornhoff GM, Shibles RM. 1970. Varietal differences in net photosynthesis of soybean leaves. *Crop Science* **10**:42–45.
- Duncan WG, McCloud DE, McGraw RL, Boote KJ. 1978. Physiological aspects of peanut yield improvement. *Crop Science* **18**:1015–1020.
- Elwell DL, Curry RB, Keener ME. 1987. Determination of potential yield-limiting factors of soybeans using SOYMOD/OARDC. *Agricultural Systems* **24**:221–242.
- Furbank RT, Tester M. 2011. Phenomics—technologies to relieve the phenotyping bottleneck. *Trends in Plant Science* **16**:635–644.
- Gilbert ME, Holbrook NM, Zwieniecki MA, Sadok W, Sinclair TR. 2011. Field confirmation of genetic variation in soybean transpiration response to vapor pressure deficit and photosynthetic compensation. *Field Crops Research* **124**:85–92.
- Grant RF, Kimball BA, Pinter PJ, Wall GW, Garcia RL, LaMorte RL, Hunsaker DJ. 1995. Carbon dioxide effects on crop energy balance: testing ecosys with a free-air CO<sub>2</sub> enrichment (FACE) experiment. *Agronomy Journal* **87**:446–457.
- Gupta SK, Rai KN, Singh P, Ameta VL, Gupta SK, Jayalekha AK, Mahala RS, Pareek S, Swami ML, Verma YS. 2015. Seed set variability under high temperatures during flowering period in pearl



- millet (*Pennisetum glaucum* L. (R.) Br.). *Field Crop Research* **171**:41–53.
- Hammer GL, Butler DG, Muchow RC, Meinke H. 1996. Integrating physiological understanding and plant breeding via crop modeling and optimization. In: Cooper M, Hammer GL, eds. *Plant adaptation and genetic improvement*. New York: CAB International, 419–441.
- Hammer GL, Chapman S, van Oosterom E, Podlich DW. 2005. Trait physiology and crop modeling as a framework to link phenotypic complexity to underlying genetic systems. *Australian Journal of Agricultural Research* **56**:947–960.
- Hammer GL, Jordan DR. 2007. An integrated systems approach to crop improvement. In: Spiertz JHJ, Struik PC, van Laar HH, eds. *Scale and complexity in plant systems research: gene-plant-crop relations*. Springer, 45–61.
- Hammer G, McLean G, van Oosterom E, Chapman S, Zheng B, Wu A, Doherty A, Jordan D. 2020. Designing crops for adaptation to the drought and high-temperature risks anticipated in future climates. *Crop Science* **60**:605–621.
- Hammer G, Messina C, van Oosterom E, Chapman S, Singh V, Borrell A, Jordan D, Cooper M. 2016. Molecular breeding for complex adaptive traits: how integrating crop ecophysiology and modelling can enhance efficiency. In: Yin X, Struik PC, eds. *Crop systems biology: narrowing the gaps between crop modelling and genetics*. Cham, Switzerland: Springer International Publishing, 147–162.
- Hammer GL, Sinclair TR, Chapman SC, Van Oosterom E. 2004. On systems thinking, systems biology, and the *in silico* plant. *Plant Physiology* **134**:909–911.
- Hammer GL, van Oosterom E, McLean G, Chapman SC, Broad I, Harland P, Muchow RC. 2010. Adapting APSIM to model the physiology and genetics of complex adaptive traits in field crops. *Journal of Experimental Botany* **61**:2185–2202.
- Hickey LT, Hafeez AN, Robinson H, Jackson SA, Leal-Bertioli SCM, Tester M, Gao C, Godwin ID, Hayes BJ, Wulff BBH. 2019. Breeding crops to feed 10 billion. *Nature Biotechnology* **37**:744–754.
- Hoogenboom G, Jones JW, Boote KJ. 1992. Modeling growth, development and yield of grain legumes using SOYGRO, PNUTGRO, and BEANGRO: a review. *Transactions of the ASAE* **35**:2043–2056.
- Hoogenboom G, Porter CH, Shelia V, Wilkens PW, Singh U, White JW, Asseng S, Lizaso JI, Moreno LP, Pavan W, Ogoshi R, Hunt LA, Tsuji GY, Jones JW. 2019a. The DSSAT crop modeling ecosystem. In: Boote KJ, ed. *Advances in crop modelling for a sustainable agriculture*. Cambridge, UK: Burleigh Dodds Science Publishing, 173–216.
- Hoogenboom G, Porter CH, Shelia V, Boote KJ, Singh U, White JW, Hunt LA, Ogoshi R, Lizaso JI, Koo J, Asseng S, Singels A, Moreno LP, Jones JW. 2019b. *Decision support system for agrotechnology transfer (DSSAT) version 4.7 (www.DSSAT.net)*. Gainesville, FL: DSSAT Foundation.
- Hoogenboom G, White JW. 2003. Improving physiological assumptions of simulation models by using gene-based approaches. *Agronomy Journal* **95**:82–89.
- Hoogenboom G, White JW, Acosta Gallegos J, Gaudiel RG, Myers JR, Silbernagel MJ. 1997. Evaluation of a crop simulation model that incorporates gene action. *Agronomy Journal* **89**:613–620.
- Hoogenboom G, White JW, Jones JW, Boote KJ. 1994. BEANGRO, a process oriented dry bean model with a versatile user interface. *Agronomy Journal* **86**:182–190.
- Hoogenboom G, White JW, Messina CD. 2004. From genome to crop: integration through simulation modeling. *Field Crops Research* **90**:145–163.
- Hwang C, Correll MJ, Gezan SA, Zhang L, Bhakta MS, Vallejos CE, Boote KJ, Clavijo-Michelangeli JA, Jones JW. 2017. Next generation crop models: a modular approach to model early vegetative and reproductive development of the common bean (*Phaseolus vulgaris* L.). *Agricultural Systems* **155**:225–239.
- Jones JW, Hoogenboom G, Porter CH, Boote KJ, Batchelor WD, Hunt LA, Wilkens PW, Singh U, Gijsman AJ, Ritchie JT. 2003. The DSSAT cropping system model. *European Journal of Agronomy* **18**:235–265.
- Kaspar TC, Stanley CD, Taylor HM. 1978. Soybean root growth during the reproductive stages of development. *Agronomy Journal* **70**:1105–1107.
- Kaspar TC, Taylor HM, Shibles RM. 1984. Taproot-elongation rates of soybean cultivars in the glasshouse and their relation to field rooting depth. *Crop Science* **24**:916–920.
- Landivar JA, Baker DN, Jenkins JN. 1983. Application of GOSSYM to genetic feasibility studies. II. Analyses of increasing photosynthesis, specific leaf weight and longevity of leaves in cotton. *Crop Science* **23**:504–510.
- Leuning R. 1995. A critical appraisal of a combined stomatal-photosynthesis model for C-3 plants. *Plant, Cell & Environment* **18**:339–355.
- Lizaso JI, Batchelor WD, Boote KJ, Westgate ME. 2005. Development of a leaf-level canopy assimilation model for CERES-Maize. *Agronomy Journal* **97**:722–733.
- Lizaso JI, Batchelor WD, Westgate ME. 2003. A leaf area model to simulate cultivar-specific expansion and senescence of maize leaves. *Field Crops Research* **80**:1–17.
- Löffler CM, Wei J, Fast T, Gogerty J, Langton S, Bergman M, Merrill B, Cooper M. 2005. Classification of maize environments using crop simulation and geographic information system. *Crop Science* **45**:1708–1716.
- Loison R, Audebert A, Debaeke P, Hoogenboom G, Leroux L, Oumarou P, Gérardaux E. 2017. Designing cotton ideotypes for the future: reducing risk of crop failure for low input rainfed conditions in Northern Cameroon. *European Journal of Agronomy* **90**:162–173.
- Messina CD, Jones JW, Boote KJ, Vallejos CE. 2006. A gene-based model to simulate soybean development and yield responses to environment. *Crop Science* **46**:456–466.
- Messina CD, Sinclair TR, Hammer GL, Curan D, Thompson J, Oler Z, Gho C, Cooper M. 2015. Limited-transpiration trait may increase maize drought tolerance in the US Corn Belt. *Agronomy Journal* **107**:1978–1986.
- Messina CD, Technow F, Tang T, Totir R, Gho C, Cooper M. 2018. Leveraging biological insight and environmental variation to improve phenotypic prediction: integrating crop growth models (CGM) with whole genome prediction (WGP). *European Journal of Agronomy*. doi:10.1016/j.eja.2018.01.007.
- Miner GL, Bauerle WL, Baldocchi DD. 2017. Estimating the sensitivity of stomatal conductance to photosynthesis: a review. *Plant, Cell & Environment* **40**:1214–1238.

- Morrison MJ, Voldeng HD, Cober ER. 1999. Physiological changes from 58 years of genetic improvement of short-season soybean cultivars in Canada. *Agronomy Journal* **91**:685–689.
- Muller B, Martre P. 2019. Plant and crop simulation models: powerful tools to link physiology, genetics, and phenomics. *Journal of Experimental Botany* **70**:2339–2344.
- Naab JB, Singh P, Boote KJ, Jones JW, Marfo KO. 2004. Using the CROPGRO-peanut model to quantify yield gaps of peanut in the Guinean savanna zone of Ghana. *Agronomy Journal* **96**:1231–1242.
- Naab JB, Tsigbey FK, Prasad PVV, Boote KJ, Bailey JE, Brandenburg RL. 2005. Effects of sowing date and fungicide application on yield of early and late maturing peanut cultivars grown under rainfed conditions in Ghana. *Crop Protection* **24**:325–332.
- Naevae SL, Miller-Garvin J. 2019. *US Soybean Quality - 2019 Annual Report*. Saint Paul, MN: University of Minnesota.
- Narh S, Boote KJ, Naab JB, Abudulai M, M'Bi Bertin Z, Sankara P, Burow MD, Tillman BL, Brandenburg RL, Jordan DL. 2014. Yield improvement and genotype x environment analyses of peanut cultivars in multilocation trials in West Africa. *Crop Science* **54**:2413–2422.
- Narh S, Boote KJ, Naab JB, Jones JW, Tillman BL, Abudulai M, Sankara P, Bertin ZM, Burow MD, Brandenburg RL, Jordan DL. 2015. Genetic improvement of peanut cultivars for West Africa evaluated with the CSM-CROPGRO-Peanut Model. *Agronomy Journal* **106**:2213–2229.
- Oliveira FAA, Jones JW, Pavan W, Bhakta M, Vallejos CE, Correll MJ, Boote KJ, Fernandes JMC, Hölb CA, Hoogenboom G. 2021. Incorporating a dynamic gene-based process module into a crop simulation model. *Special Issue in In Silico Plants*. doi:[10.1093/inilicoplants/diab0011](https://doi.org/10.1093/inilicoplants/diab0011).
- Parent B, Tardieu F. 2014. Can current crop models be used in the phenotyping era for predicting the genetic variability of yield of plants subjected to drought or high temperature. *Journal of Experimental Botany* **65**:6179–6189.
- Peng S, Khush GS, Virk P, Tang Q, Zou Y. 2008. Progress in ideotype breeding to increase rice yield potential. *Field Crops Research* **108**:32–38.
- Penning deVries FWT, van Laar HH. 1982. Simulation of growth processes and the model BACROS. In: Penning deVries FWT, van Laar HH, eds. *Simulation of plant growth and crop production*. Wageningen, The Netherlands: Pudoc, 114–135.
- Pickering NB, Jones JW, Boote KJ. 1995. Adapting SOYGRO V5.42 for prediction under climate change conditions. In: Rosenzweig C, Jones JW, Allen LH Jr, eds. *Climate change and agriculture: analysis of potential international impacts*, ASA Spec. Pub. No. 59. Madison, WI: ASA-CSSA-SSSA, 77–98.
- Prasad PVV, Boote KJ, Allen, LH Jr, Thomas JM. 2003. Supra-optimal temperatures are detrimental to peanut (*Arachis hypogaea* L.) reproductive processes and yield at ambient and elevated carbon dioxide. *Global Change Biology* **9**:1775–1787.
- Putto C, Pathanothai A, Jogloy S, Boote KJ, Hoogenboom G. 2013. Determination of plant traits that affect genotype x location (G x L) interaction in peanut using the CSM-CROPGRO-Peanut model. *International Journal of Plant Production* **7**:537–568.
- Putto C, Pathanothai A, Jogloy S, Hoogenboom G. 2008. Determination of mega environments for peanut breeding using the CSM CROPGRO Peanut model. *Crop Science* **48**:973–982.
- Reymond MB, Muller B, Leonardi A, Charcosset A, Tardieu F. 2003. Combining quantitative trait loci analysis and an ecophysiological model to analyze the genetic variability of the responses of maize leaf growth to temperature and water deficit. *Plant Physiology* **131**:664–675.
- Ritchie JT. 1998. Soil water balance and plant water stress. In: Tsuji GY, Hoogenboom G, Thornton PK, eds. *Understanding options for agricultural production*. Dordrecht, The Netherlands: Kluwer Academic Publishers, 41–54.
- Ritchie JT, Singh U, Godwin DC, Bowen WT. 1998. Cereal growth, development and yield. In: Tsuji GY, Hoogenboom G, Thornton PK, eds. *Understanding options for agricultural production*. Dordrecht, The Netherlands: Kluwer Academic Publishers, 79–98.
- Sinclair TR, Messina CD, Beatty A, Samples M. 2010. Assessment across the United States of the benefits of altered soybean drought traits. *Agronomy Journal* **102**:475–482.
- Singh P, Alagarswamy G, Hoogenboom G, Pathak P, Wani SP, Virmani SM. 1999a. Soybean-chickpea rotation on Vertic Inceptisols: 2. Long-term simulation of water balance and crop yields. *Field Crops Research* **63**:225–236.
- Singh P, Alagarswamy G, Pathak P, Wani SP, Hoogenboom G, Virmani SM. 1999b. Soybean-chickpea rotation on Vertic Inceptisols: 1. Effect of soil depth and landform on light interception, water balance and crop yields. *Field Crops Research* **63**:211–224.
- Singh P, Boote KJ, Kadiyala MDM, Nedumaran S, Gupta SK, Srinivas K, Bantilan MCS. 2017. An assessment of yield gains under climate change due to genetic modification of pearl millet. *Science of the Total Environment* **601–602**:1226–1237.
- Singh P, Boote KJ, Kumar U, Srinivas K, Nigam SN, Jones JW. 2012. Evaluation of genetic traits for improving productivity and adaptation of groundnut to climate change in India. *Journal of Agronomy and Crop Science* **198**:399–413.
- Singh P, Nedumaran S, Boote KJ, Gaur PM, Srinivas K, Bantilan MCS. 2014. Climate change impacts and potential benefits of drought and heat tolerance in chickpea in South Asia and East Africa. *European Journal of Agronomy* **52**:123–137.
- Singh P, Virmani SM. 1994. Modeling growth and yield of chickpea (*Cicer arietinum* L.). *Field Crops Research* **46**:1–29.
- Surihar B, Patanothai A, Boote KJ, Hoogenboom G. 2011. Designing a peanut ideotype for a target environment using the CSM CROPGRO Peanut model. *Crop Science* **51**:1887–1902.
- Surihar B, Patanothai A, Pannangpetch K, Jogloy S, Hoogenboom G. 2007. Determination of cultivar coefficients of peanut lines for breeding applications of the CSM-CROPGRO-Peanut model. *Crop Science* **47**:607–619.
- Tardieu F. 2003. Virtual plants: modelling as a tool for the genomics of tolerance to water deficit. *Trends in Plant Science* **8**:9–14.
- Tardieu F. 2012. Any trait or trait-related allele can confer drought tolerance: just design the right drought scenario. *Journal of Experimental Botany* **63**:25–31.
- Tardieu F, Cabrera-Bosquet L, Pridmore T, Bennett M. 2017. Plant phenomics, from sensors to knowledge. *Current Biology* **27**:770–783.
- Technow F, Messina CD, Totir LR, Cooper M. 2015. Integrating crop growth models with whole genome prediction through approximate Bayesian computation. *PLoS One* **10**:e0130855.
- Thorburn PJ, Boote KJ, Nendel C, Rötter RP, Ewert F. 2018. Recent advances in crop modelling to support sustainable agricultural

- production and food security under global change. *European Journal of Agronomy* **100**:1–3.
- Tsuji GY, Hoogenboom G, Thornton PK (eds). 1998. *Understanding options for agricultural production*. Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Varshney RK, Sinha P, Singh VK, Kumar A, Zhang Q, Bennetzen JL. 2020. 5Gs for crop genetic improvement. *Current Opinion in Plant Biology* **56**:190–196.
- White JW, Herndl M, Hunt LA, Payne TS, Hoogenboom G. 2008. Simulation-based analysis of effects of *Vrn* and *Ppd* loci on flowering in wheat. *Crop Science* **48**:678–687.
- White JW, Hoogenboom G. 1996. Simulating effects of genes for physiological traits in a process-oriented crop model. *Agronomy Journal* **88**:416–422.
- White JW, Hoogenboom G. 2003. Gene-based approaches to crop simulation: past experiences and future opportunities. *Agronomy Journal* **95**:52–64.
- Yin X, Kropff MJ, Stam P. 1999. The role of ecophysiological models in QTL analysis: the example of specific leaf area in barley. *Heredity* **82**:415–421.
- Yin X, Struik P (eds). 2016. *Crop systems biology. Narrowing the gaps between crop modeling and genetics*. Cham, Switzerland: Springer International Publishing.
- Yin X, Struik PC, Kropff MJ. 2004. Role of crop physiology in predicting gene-to-phenotype relationships. *Trends in Plant Science* **9**:426–432.
- Yin X, Struik PC, Tang J, Qi C, Liu T. 2005a. QTL analysis and QTL-based prediction of flowering phenology in recombinant inbred lines of barley. *Journal of Experimental Botany* **56**:967–976, 2003.
- Yin X, Struik PC, Tang J, Qi C, Liu T. 2005b. Model analysis of flowering phenology in recombinant inbred lines of barley. *Journal of Experimental Botany* **56**:959–965.
- Zheng B, Biddulph B, Li D, Kuchel H, Chapman S. 2013. Quantification of the effects of *VRN1* and *Ppd-D1* to predict spring wheat (*Triticum aestivum*) heading time across diverse environments. *Journal of Experimental Botany* **64**:3747–3761.
- Zheng B, Chenu K, Chapman SC. 2016. Velocity of temperature and flowering time in wheat – assisting breeders to keep pace with climate change. *Global Change Biology* **22**:921–933.