

ORIGINAL ARTICLE

Crop Breeding & Genetics

Calibration of a crop growth model in APSIM for 15 publicly available corn hybrids in North America

Cassandra Anne Winn¹ | Sotirios Archontoulis²  | Jode Edwards³ 

¹Benson Hill Inc., St. Louis, MO, USA

²Department of Agronomy, Iowa State University, Ames, IA, USA

³USDA-ARS, Corn Insects and Crop Genetics Research Unit, Ames, IA, USA

Correspondence

Jode Edwards, USDA-ARS, Corn Insects and Crop Genetics Research Unit, 716 Farmhouse Ln., Ames, IA 50011, USA.
Email: jode.edwards@usda.gov

Assigned to Associate Editor Sivakumar Sukumaran.

Funding information

Foundation for Food and Agriculture Research, Grant/Award Number: 534264; Division of Civil, Mechanical and Manufacturing Innovation, Grant/Award Number: 1830478; Division of Integrative Organismal Systems, Grant/Award Number: 1842097; Agricultural Research Service, Grant/Award Number: 5030-21000-066-00D; Cooperative State Research, Education, and Extension Service, Grant/Award Numbers: IOW04614, IOW10480; Division of Graduate Education, Grant/Award Number: 1545453; National Institute of Food and Agriculture, Grant/Award Number: 2017-67013-26463

Abstract

Application of crop growth models (CGMs) in plant breeding is limited by the large number of candidate cultivars that breeders work with and the large number of CGM parameters that affect cultivar performance. The objectives of this study were to (1) calibrate 15 publicly available maize hybrids in Agricultural Production Systems sIMulator and quantify prediction accuracy in modeling physiological trait differences (yield, biomass, phenology, etc.) among genotypes; (2) better understand minimum phenotypic data requirements for CGM cultivar calibration to inform breeding efforts; and (3) quantify simulated genotype by environment interactions ($G \times E$) across years for five traits. We calibrated hybrids with two years of multi-trait, temporal field measurements. The R^2 of simulated versus observed phenotypes was 0.89 for grain yield and over 0.80 for half of all other simulated traits. Phenology parameters accounted for nearly half of the variability in grain yield. Average (across traits) normalized root mean square error was reduced from 35% to 30% with calibration based on phenological measurements and was reduced to 20% with inclusion of physiological and nitrogen-related measurements such as radiation use efficiency and grain nitrogen. Long-term simulations demonstrated distinct $G \times E$ among the hybrids which accounted for 2%–29% of the total genetic variation across traits. Parameter values derived in this work will provide insight regarding important physiological traits for further phenotyping, selection, and understanding of $G \times E$. These calibrations are for publicly available hybrids, which are currently lacking.

Abbreviations: APSIM, Agricultural Production Systems sIMulator; CGM, crop growth model; ex-PVP, expired plant variety protection certificate; G2F, Genomes to Fields; $G \times E$, genotype by environment interaction; LAI, leaf area index; NRMSE, normalized root mean square error; RUE, radiation use efficiency.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. Crop Science © 2022 Crop Science Society of America.

1 | INTRODUCTION

The need to increase yields by 2050 to meet predicted global food production demand (Godfray et al., 2010; Mueller et al., 2012; Ray et al., 2013) warrants new tools and methods to improve agronomics and breeding. The goal of plant breeders is to develop and select high performing, stable varieties for specific target environments. Genotype by environment

interactions ($G \times E$) is common, making such a goal difficult to attain. Therefore, the prediction of $G \times E$ is of practical importance for improved selection of high performing cultivars across highly variable environments. Extensive work has been done to develop statistical, descriptive models and methods for evaluating $G \times E$. Studies have explored adoption of genomic selection models for use in predicting $G \times E$ by attributing environment-specific effects to the markers (Crossa et al., 2016; Schulz-Streeck et al., 2013), introducing environmental covariates (Heslot et al., 2014; Jarquin et al., 2014; Malosetti et al., 2016), or modeling environmental covariances (Burgueno et al., 2012). In most cases, these approaches reduce $G \times E$ to linear relationships between varieties and only a few major environmental covariates, not allowing for complex interactions. These linear statistical models limit prediction to environments that are similar to those already observed.

Dynamic, process-based crop growth models (CGMs) provide a framework for understanding $G \times E$ with the incorporation of biological knowledge, thereby offering interpretive and predictive potential beyond purely descriptive statistical models (Hammer et al., 2006; Technow et al., 2015). Prediction of environmental-specific performance is challenged by the biological complexities associated with genetic factors, environmental effects, agronomic practices, and their interactions (Hammer et al., 2006). CGMs attempt to unravel such complexity and provide an opportunity to explore, through simulation, the processes underlying the interactions and the consequences associated with breeding and agronomics (Chenu et al., 2017; Hammer et al., 2014). In CGMs, plant and soil knowledge are integrated using a series of equations and coefficients to quantify elements known to be variable across cultivars and soils.

Historically, physiological knowledge was used to provide explanations for improvement already achieved from direct selection on yield as opposed to guiding future crop improvement (Donald, 1968). Nearly 30 years ago, Jackson et al. (1996) and Shorter et al. (1991) highlighted the need for collaborative efforts among crop physiologists and plant breeders while using crop models as a basis for integration. Since then, successful applications of CGMs in plant breeding have been demonstrated including designing multiple trait ideotypes (Dingkuhn et al., 1993; Haverkort & Kooman, 1997), evaluating breeding strategies (Chapman et al., 2003; Messina et al., 2011), characterizing $G \times E$ by understanding how the interactions arise during crop growth and development (Bustos-Korts et al., 2019), predicting observed $G \times E$ through imputation of genetic parameters (Mavromatis et al., 2001), evaluating phenotyping strategies and the incorporation of component trait data to improve genomic prediction (Bustos-Korts et al., 2019; van Eeuwijk et al., 2019), characterizing environments relative to stress (Chapman et al., 2000; Löffler et al., 2005), understanding crop production

Core Ideas

- Calibration of 15 public maize hybrids within APSIM using multi-trait time series data.
- Compared to no calibration, phenology data reduced average across-trait and hybrid NRMSE from 35% to 30% and crop growth data reduced it to 20%.
- APSIM simulated $G \times E$ proportions of variance and trends observed by breeders for several traits across 20 years.

implications of various management strategies (Kheir et al., 2021), and integrating CGMs with whole-genome prediction (Cooper et al., 2016; Messina et al., 2018; Technow et al., 2015). Recent studies also suggest the use of detailed genomic information and a well-structured CGM to statistically impute key parameters, facilitating CGM calibration at the scale of a breeding population (Cooper et al., 2016; Cooper, Messina, et al., 2014; Messina et al., 2018). Imputation approaches alleviate the need for large amounts of extensive phenotypic data for cultivar-specific parameter calibration, but put added pressure on the structure of the CGM to be both biologically sound and parsimonious (Hammer et al., 2019). The benefit CGMs can provide to plant breeding largely depends on the structure of the CGM and whether the CGM algorithms appropriately capture the physiological determinants underlying genetic variation for traits of interest, such as yield (Cooper et al., 2016, 2009; Hammer et al., 2019).

Existing applications of CGMs in plant breeding, as well as many management-oriented uses of CGMs, rely heavily on generic cultivars and a simulated subset of hybrid-specific model parameters. In many approaches, modelers have grouped cultivars by maturity zones and calibrated generic, modern cultivars with many presumed parameters for use across environments (Archontoulis et al., 2014; Boote et al., 2003; Yang et al., 2004). This approach ignores many sources of physiological variation among cultivars beyond phenology. Other studies have gone a step further and included subsets of specific parameters as latent variables, thereby leveraging the knowledge embedded in the CGM to understand phenotypes that are difficult to measure directly (Cooper et al., 2016; Messina et al., 2018; Technow et al., 2015). However, few publicly available, comprehensive cultivar calibrations exist. Empirical calibration of cultivars requires intense phenotyping to estimate large numbers of CGM parameters intertwined in complex processes affecting cultivar performance. Relating the phenotypic data to such CGM parameters is another challenge (Cooper, Gho, et al., 2014; Hammer et al., 2019; Messina et al., 2011, 2018;

van Eeuwijk et al., 2019). The large number of parameters and the labor required limits the number of cultivars for which a complete calibration can be obtained, explaining why few comprehensive calibrations exist. Cultivar calibrations would enable additional applications and improve connections among modelers, physiologists, and plant breeders (Hammer et al., 2010; Struik et al., 2007). Future advances in high-throughput phenomics (Reynolds et al., 2020) offer the potential to generate increased quantity and quality of data to improve cultivar-specific calibration (Furbank & Tester, 2011). Nevertheless, the lack of empirical, fully-calibrated cultivars creates potential gaps in our ability to test model structure and rigor with respect to all cultivar-specific parameters. Detailed, empirical calibrations will enable evaluation of the ability and limitations of CGMs to accurately simulate observed differences among modern cultivars and predict $G \times E$.

A step toward improving connections between plant breeding and crop modeling and assessing the CGM structural validity for fitting across genotypes and environments is to collect comprehensive, precise phenotypes and subsequently calibrate a CGM for extensively studied public hybrids. This approach will familiarize plant breeders with the data requirements and potential benefits of CGMs, while testing the model fit accuracy and ability to simulate plausible, expected phenotypic responses across genotypes (Hammer, 2020). Developing parameter values for publicly available ex-PVP (expired Plant Variety Protection certificate) maize hybrids is significant because genotypic marker data and pedigree information are readily available and could be used to establish relationships between underlying genetics and cultivar parameters. Despite no longer being used by farmers today, public-sector institutions rely on public hybrids for research purposes and have generated large amount of data from genetic diversity analyses and genome-wide association studies (Beckett et al., 2017; Kusmec et al., 2021; Romay et al., 2013). Public hybrids are often used in collaborative, multi-institutional trials such as those conducted under the Genomes to Fields Initiative (G2F; genomes2fields.org), providing open-access to phenotypic data across diverse environments (McFarland et al., 2020). Currently, trait data collected in most plant breeding trials (grain yield, grain moisture, plant height, lodging, and flowering date) are not enough to calibrate hybrids for use in CGMs (Archontoulis et al., 2020; Cooper et al., 2016).

Using the Agricultural Production Systems sIMulator (APSIM) maize, a model recently upgraded with additional physiological mechanisms to accommodate breeding research (Hammer et al., 2010; Soufizadeh et al., 2018), our objective was to develop parameter values for ex-PVP hybrids for which genotypic data are available. Our specific objectives were to (1) calibrate APSIM for 15 publicly available maize hybrids

TABLE 1 Number of locations each hybrid was grown and field data collected

Hybrid	2017	2018
B73 × Mo17	1	2
B73 × PHM49	1	2
B73 × PHZ51	1	2
B73 × LH185	0	2
LH195 × Mo17	0	2
LH195 × PHM49	1	2
LH195 × PHZ51	1	2
LH195 × LH185	1	2
PHW52 × Mo17	0	2
PHW52 × PHM49	1	2
PHW52 × PHZ51	1	2
PHW52 × LH185	1	2
PHJ40 × LH82	1	0
PHZ51 × LH145	1	0
LH145 × LH162	1	0

and quantify prediction accuracy in modeling physiological trait differences among genotypes; (2) better understand minimum data requirements for accurate model calibration to inform breeding efforts; and (3) quantify $G \times E$ across many years for five traits: grain yield, grain nitrogen (N) concentration, harvest index, maximum leaf area index (LAI_{max}), and flowering time.

2 | MATERIALS AND METHODS

2.1 | Maize hybrids

The 15 hybrids used in this study were assembled from 11 ex-PVP inbreds developed and grown in the U.S. Corn Belt (Table S1). Twelve of the hybrids were chosen because they are grown widely in the G2F project, providing access to performance data from a range of 21–136 year × location environments across the United States (Anderson et al., 2019; Gage et al., 2017; Lawrence-Dill et al., 2019; McFarland et al., 2020; Sekhon et al., 2020). A subset of the 15 hybrids constituted a factorial mating design of three females: B73, LH195, and PHW52, and four males: PHZ51, LH185, PHM49, and Mo17. Twelve of the 15 hybrids were grown in the first year (2017), and a modified set of 12 hybrids were grown in the second year (2018; Table 1). One hybrid in particular, B73 × Mo17, was included in the study because B73 is an important public founder line (Schnable et al., 2009) from which many present-day commercial lines and hybrids are derived (Mikel & Dudley, 2006).

2.2 | Field experiments

We planted hybrids in four environments across two years (2017, 2018) and three locations in central Iowa, (AGR–Agronomy Farm, BRNE–Bruner Farm, and JHN–Johnson Farm), but collected data from only three experiments (AGR 2017: 42.0210, –93.7779; BRNE 2018: 42.0117, –93.7364; JHN 2018: 41.9799, –93.6457) as the JHN 2017 (41.9833, –93.6420) experiment was severely damaged by wind and discarded. Within each environment, hybrids were planted in a randomized complete block design with six replications. Three replications were used for destructive phenotyping and three replications for non-destructive phenotyping and end of season combine harvesting (see Section 2.4). Individual plots were four rows (spaced 76 cm apart) and 11.35 m long. Plots were planted at 8.8 plants m^{-2} with precision planting equipment. The 2017 AGR location was field cultivated on April 18 and planted on May 8, and 150 kg N/ha of 32% UAN-N fertilizer was applied on April 25. The 2018 BRNE location was field cultivated on April 18 and planted on May 16, and 168 kg N/ha 32% UAN-N fertilizer was applied on April 27. The 2018 JHN location was field cultivated on May 17 and planted on May 18, and 157 kg N/ha 32% UAN-N fertilizer was applied on May 17. The previous crop was soybean in all experiments.

2.3 | Weather data

In the BRNE locations, daily weather data were retrieved from Iowa Environmental Mesonet (Iowa Environmental Mesonet, 2019). In the JHN location, temperature and precipitation data were retrieved from an in-field weather station, and radiation was sourced from NASAPower (<https://power.larc.nasa.gov/>). All environments had an average summer maximum temperature of 29°C (Figure 1a). Total summer precipitation (June, July, August) ranged from 199 to 574 mm among environments with a total yearly precipitation from 331 to 715 mm (Figure 1b). In general, 2017 was dry, while 2018 was very wet with extreme rain events and periodic flooding, which inhibited root depth (Archontoulis et al., 2020; Ebrahimi-Mollabashi et al., 2019).

2.4 | Phenotyping measurements

Non-destructive measurements included the following: (1) emergence date (date when 50% of the plants emerged and were visible), (2) number of plants at approximately 6th leaf stage (Abendroth et al., 2011)—plant counts were on average 97% of the target population, (3) number of green collared leaves and number of senesced leaves (>50% leaf area yellow)

low) at each harvest date, (4) dates when 50% of the plants in the center two rows were shedding pollen and 50% of plants had visible silks, (5) physiological maturity dates determined by a fully formed black layer according to Hunter et al. (1991), and (6) number of plants stalk and root lodged in the center two rows prior to harvest.

Destructive measurements were taken five times in 2017 and four times in 2018 from each plot. Each time, we harvested a 1-m section from one of the two center plot rows. Harvest dates were designed to capture key growth stages, namely V8 (8 collared leaves), R1 (silking), R2 (blister), R4 (dough), and R6 (physiological maturity) (Abendroth et al., 2011). At each harvest, plants were cut at ground level, and the number of harvested plants was recorded. Any detached, senesced leaves underneath the harvested plants were collected. Following harvest, harvested plants were partitioned to (1) green leaf lamina (defined by a leaf having at least 50% of total area green), (2) senesced leaves, (3) stems, (4) tassels, and (5) ears, which were further partitioned to kernels, cobs, husks, and shanks. Green leaf samples were scanned on a per-plant basis using a table-top LICOR leaf area meter to determine LAI. All plant biomass samples were dried at 60°C to constant mass, weighed, and converted to grams dry biomass per m^2 . Samples of 100 kernels were counted and weighed. Grain number was calculated using total dry weight per unit area and 100 kernel dry weight. Grain yield (derived from hand-harvested grain biomass in g/m^2 at 0% moisture) was converted to kg/ha. Dried samples were ground and analyzed for carbon (C) and N concentrations using dry combustion elemental analysis (LECO C and N analyzer; LECO Corporation, St. Joseph, Michigan). Nitrogen concentration was converted to $g N m^{-2}$ for each sample type and summed across samples to compute total N uptake. Final grain yields were also determined from the center two rows of plots designated for non-destructive phenotyping using a New Holland TR88 but were only used to compare to hand-harvested yields.

2.5 | The APSIM maize model

APSIM is a widely used cropping system model platform (Holzworth et al., 2014; Keating et al., 2003) that enables the simulation of soil-crop-environment interactions. In this study, we used APSIM version 7.10, specifically the maize crop model module (Soufizadeh et al., 2018), SWIM for simulation of soil water (Huth et al., 2012), SOILN for simulation of soil carbon and N cycling (Probert et al., 1998), and various management rules to account for tillage, fertilizer application, and planting conditions.

The maize model simulates phenology, leaf development and senescence, and biomass accumulation and partitioning across nine crop phases. The duration of each phase is

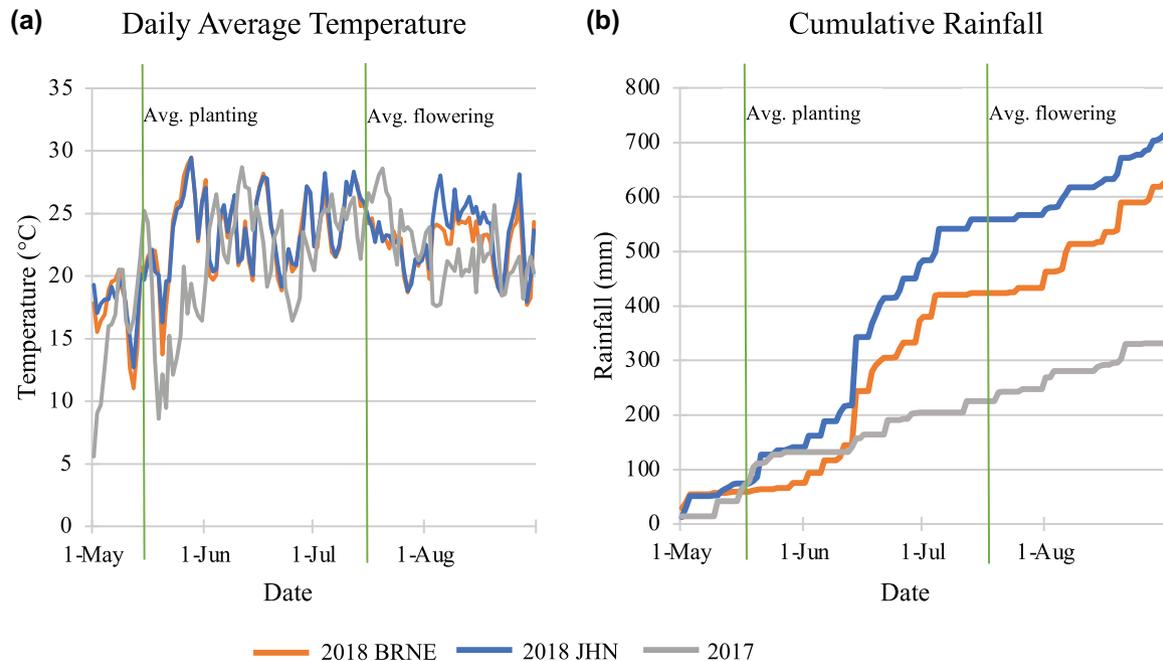


FIGURE 1 Daily average temperature ($^{\circ}\text{C}$; panel a) and cumulative rainfall (mm; panel b), at all environments during the growing season from weather data used for simulations in each environment

primarily dependent on temperature, and to a lesser extent, moisture stress and photoperiod. Canopy development is driven mainly by temperature through effects on the phyllochron, with final daily estimates of LAI also dependent on carbon availability (via specific leaf area) and water stress. Final leaf number is an emergent property of the model and is determined in part by the timing of floral initiation. Crop growth rate and biomass production are calculated as the minimum of two daily estimates, one limited by light (radiation use efficiency, RUE) and one limited by water (TE; transpiration efficiency adjusted for vapor pressure deficit). The growth of the major organs is a function of their potential growth and whether carbohydrate and N supply can meet that demand. Biomass partitioning is stage dependent. From emergence to flowering, leaves have priority over stalk, while from flowering to physiological maturity, grain is the strongest sink. Supply of carbohydrates to grain is first fulfilled by current assimilation, and then supplemented by translocation from stem and leaves. The APSIM maize model was recently updated with new physiologically sound plant N routines and grain growth dynamics (Soufizadeh et al., 2018). The model uses specific leaf N (SLN) as a key driver in simulating plant N demand. Grain yield is simulated as the product of grain number and grain size. Grain number is estimated by a function that relates potential grain number per ear and daily crop growth rate around the critical period of silking (Edmeades & Daynard, 1979), which is also affected by heat stress. There is a strong trade-off in the model structure between grain number and individual grain weight (poten-

tial grain size = potential grain mass per plant/potential grain number) (Soufizadeh et al., 2018). The demand for carbohydrate in the grain is the product of grain number and potential grain growth rate, which is based on potential grain size and duration of grain fill (Soufizadeh et al., 2018). We refer to Soufizadeh et al. (2018) for more information on the maize model, and to Holzworth et al. (2014), Keating et al. (2003), and www.apsim.info/Documentation for APSIM in general. The soil routines of APSIM have been extensively tested in this environment (Archontoulis et al., 2020; Archontoulis et al., 2014; Dietzel et al., 2016; Martinez-Feria et al., 2018; Puntel et al., 2016).

2.6 | Sensitivity analysis

To further understand the sensitivity of the maize model to different parameter values and to guide and refine our calibrations, we performed two sensitivity analyses. The first was performed prior to calibration using the commercial Pioneer hybrid (P1197; data not shown). The second was performed after calibration using the average parameter values from all 15 hybrids as a baseline (Table S3). Twenty-six crop parameters were evaluated in the sensitivity analysis and were categorized into crop development, crop growth, and grain components. Sensitivity analyses were conducted one-factor-at-a-time (Lenhart et al., 2002) by varying a single parameter within feasible minimum and maximum values (Table S3) and running a 20-year APSIM simulation for each variation.

Model outputs included grain yield, above-ground biomass, crop N uptake, grain number, grain N uptake, grain protein, grain size, harvest index, maximum LAI, and root depth (average of 20-year simulations). The model output sensitivity to each trait was assessed by calculating the relative sensitivity index $(D_{\max} - D_{\min})/D_{\max}$ (Hamby, 1994). D_{\min} and D_{\max} are minimum and maximum output values, respectively. Data analysis was completed in R version 3.6.1 (R Development Core Team, 2019) using packages tidyverse (Wickham, 2017) and readxl (Wickham & Bryan, 2019).

2.7 | APSIM model calibration

We started simulations on January 1, approximately 5 months before planting, to allow soil water balance to reach an equilibrium. This assumption of equilibrium was tested by varying the initial water as above, below, or at field capacity and demonstrating through simulation that the soil water reached equilibrium prior to planting (data not shown). Soil profile (Table S2) and initial conditions were taken from Archontoulis et al. (2020) where past experimental data were used from the same geographical area. The simulation included fluctuating shallow water tables and inhibition of root growth as described in Ebrahimi-Mollabashi et al. (2019), but not waterlogging functions (Pasley et al., 2020).

Management information was used to simulate growth of a previously calibrated commercial Pioneer hybrid (P1197) of 111-day maturity commonly grown in this region (Baum et al., 2020). This simulation provided a baseline control which we used to compare model fit improvement for the 15 ex-PVP hybrids. The second step was to calculate initial parameter values per hybrid using the 2017 experimental data. For example, thermal time from emergence to floral initiation ($^{\circ}\text{C}\cdot\text{d}$) was calculated as $(\text{maximum leaf number} - 6) \times 23.2$, where 6 is the number of leaf initials in the embryo (Birch et al., 1998) and 23.2 is leaf initiation rate (Carberry et al., 1989; Soufizadeh et al., 2018). The thermal time from emergence to floral initiation consists of two phases: thermal time from emergence to end of juvenile (photoperiod insensitive phase) and end of juvenile to floral initiation (photoperiod sensitive phase). In this study, we set the end of juvenile to floral initiation to zero because hybrids are not sensitive to photoperiod in this region. Thermal time from flowering to start of grain fill was extrapolated from grain growth time-series observations. Root depth parameter values were based on information derived from prior studies in this environment (Ordonez et al., 2018). Next, we performed iterative manual calibration for each hybrid by changing model parameters in the following order: phenology, biomass production and leaf area, biomass partitioning, and N-related parameters (Figure 2). Calibration efforts were also first focused on parameters that are defined in APSIM to be cultivar specific

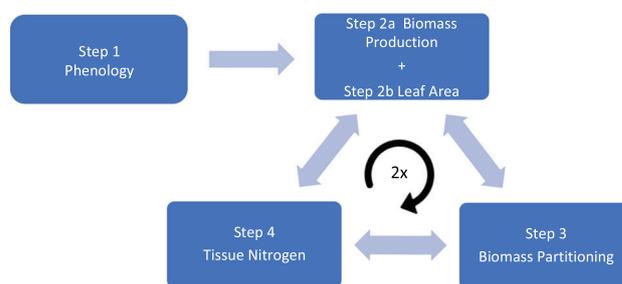


FIGURE 2 Calibration flow diagram. After reaching step 4, we looped around two times to ensure the model was producing sensible results relating to biomass production (step 2a), leaf area (step 2b), biomass partitioning (step 3), and tissue nitrogen (step 4)

and subsequently expanded to parameters not categorized as cultivar specific, but identified as important in the initial sensitivity analyses and known to vary among hybrids, such as leaf appearance rates (Hebert, 1990; Tollenaar et al., 1984), RUE (Tollenaar & Aguilera, 1992), target grain N concentration (DeBruin et al., 2017; Yan et al., 2014; Zhang et al., 2020), and critical specific leaf N (Ordonez et al., 2015). The model has many feedbacks between processes, and in some cases, a change in a morphological parameter (e.g., specific leaf area) had effects on N variables as well. Another example of this is in simulated grain number. Grain number is sensitive to a parameter representing the tradeoff relationship between potential grain number and grain size, which also influences genetic yield potential (Table 2). The ex-PVP hybrids used in this study have reduced yield potential compared to many of today's commercial hybrids and a range of observed grain numbers and grain sizes that warranted inclusion of this parameter in the calibration process. The parameter values were manually adjusted using knowledge of maize crop growth and development and visual examination of plots comparing observed and simulated traits across the growing season (see Figure 5). Our goal in calibration was for the model to adequately predict all phenotypes, not just yield. The first set of hybrid calibrations, attained using only 2017 data (one environment) and prior knowledge, will be referred to in this study as "calibration set one." We tested calibration set one in simulating 2018 experimental data to evaluate overall model performance. In an attempt to better simulate what was observed in the waterlogged 2018 environments, we performed an ad hoc change in APSIM so that when the simulated water table was less than 30 cm from the soil surface, we reduced solar radiation by 90% for that period (usually 1–3 days), in an effort to reproduce the inhibition of plant growth caused by soil oxygen stress (Pasley et al., 2020). In the final step, we used all experimental data (three environments) and the protocol outlined above to recalibrate the model. Calibration concluded once a good balance between observed and simulated values was reached across the growing season in all three environments for traits including grain

TABLE 2 Calibrated maize parameter values (calibration set two) for 15 public hybrids

Hybrid	Parameter																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
B73 × Mo17	310	700	170	290	165	1.85	62	36	675	26,000, 8000	1.97	0.0091	0.34	0.01	0.05	0.05, 0.05, 0.01, 0.005	0, 7.5, 15, 25, 30, 30, 15, 0, 0, 0, 0, 0
B73 × PHM49	315	800	120	320	185	1.8	62	36	900	26,000, 8000	1.97	0.0091	0.34	0.01	0.01	0.04, 0.04, 0.01, 0.005	0, 7.5, 15, 25, 30, 30, 15, 0, 0, 0, 0, 0
B73 × PHZ51	315	690	75	310	185	1.85	65	36	700	26,000, 8000	1.97	0.0182	0.34	0.01	0.05	0.05, 0.05, 0.01, 0.005	0, 7.5, 15, 25, 30, 30, 15, 0, 0, 0, 0, 0
LH145 × LH162	250	660	120	235	130	1.4	65	30	800	26,000, 8000	1.97	0.0182	0.34	0.01	0.01	0.05, 0.05, 0.01, 0.005	0, 7.5, 15, 25, 30, 30, 30, 0, 0, 0, 0, 0
LH195 × LH185	310	760	110	265	140	1.75	65	36	879	26,000, 8000	1.7	0.0091	0.34	0.01	0.05	0.06, 0.06, 0.01, 0.005	0, 7.5, 15, 25, 30, 30, 30, 0, 0, 0, 0, 0
LH195 × PHM49	320	750	110	285	170	1.85	65	36	879	26,000, 8000	1.7	0.0091	0.34	0.01	0.05	0.04, 0.04, 0.01, 0.005	0, 7.5, 15, 25, 30, 30, 15, 0, 0, 0, 0, 0
LH195 × PHZ51	310	770	130	250	150	1.8	65	38	840	26,000, 8000	1.9	0.0091	0.34	0.0085	0.05	0.06, 0.06, 0.01, 0.005	0, 7.5, 15, 25, 30, 30, 15, 0, 0, 0, 0, 0
PH140 × LH82	250	775	110	295	155	1.75	65	36	700	26,000, 8000	1.7	0.0182	0.34	0.009	0.05	0.06, 0.06, 0.01, 0.005	0, 7.5, 15, 25, 30, 30, 30, 0, 0, 0, 0, 0
PHW52 × LH185	285	820	130	305	145	1.65	65	36	750	26,000, 8000	1.97	0.0091	0.34	0.01	0.05	0.05, 0.05, 0.01, 0.005	0, 7.5, 15, 25, 30, 30, 15, 0, 0, 0, 0, 0
PHW52 × PHM49	300	818	100	300	165	1.75	65	36	690	23,000, 8000	1.7	0.0091	0.34	0.01	0.05	0.04, 0.04, 0.01, 0.005	0, 7.5, 15, 25, 30, 30, 15, 0, 0, 0, 0, 0

(Continues)

TABLE 2 (Continued)

Hybrid	Parameter																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
PHW52 × PHZ51	300	750	110	290	170	1.85	65	36	800	26,000,	1.7	0.0091	0.34	0.0085	0.05	0.04, 0.04,	0, 7.5, 15, 25, 30,
									8000							0.01, 0.005	30, 15, 0, 0, 0,
																	0, 0
PHZ51 × LH145	270	840	180	265	150	1.65	65	36	900	26,000,	1.9	0.012	0.34	0.009	0.05	0.06, 0.06,	0, 7.5, 15, 25, 30,
									8000							0.01, 0.005	30, 30, 0, 0, 0,
																	0, 0
LH195 × Mo17	275	770	200	240	120	1.8	65	36	900	26,000,	1.7	0.0091	0.34	0.01	0.05	0.04, 0.04,	0, 7.5, 15, 25, 30,
									8000							0.01, 0.005	30, 15, 0, 0, 0,
																	0, 0
PHW52 × Mo17	280	740	175	260	135	1.85	65	36	523	26,000,	1.7	0.0091	0.34	0.0085	0.05	0.04, 0.04,	0, 7.5, 15, 25, 30,
									8000							0.01, 0.005	30, 15, 0, 0, 0,
																	0, 0
B73 × LH185	250	770	170	260	125	1.85	65	32	523	26,000,	1.7	0.0091	0.34	0.0085	0.05	0.04, 0.04,	0, 7.5, 15, 25, 30,
									8000							0.01, 0.005	30, 30, 0, 0, 0,
																	0, 0
Default Value	280	850	170	320	210	1.85	65	36	683	8000,	1.7	0.0182	0.17	0.01	0.01	0.04, 0.04,	0, 7.5, 15, 25, 30,
									8000							0.01, 0.005	30, 30, 0, 0, 0,
																	0, 0
Average	289	761	134	278	153	1.76	64.6	35.3	784	–	1.8	0.011	0.34	0.009	0.045	–	–
Standard deviation	25	49	36	26	21	0.12	1.1	1.8	138.3	–	0.13	0.003	0	0.0007	0.014	–	–
% Dev. from mean	7.6	4.8	22.4	8.2	18.5	4.8	1.1	2.8	14.2	–	6.9	26.6	0	6.8	20.7	–	–

Notes: Each parameter number, for example, 1, corresponds to an Agricultural Production Systems sIMulator (APSIM) parameter trait—

1: $t_{emerge_to_endjuv}$ (°C-d) + t_{endjuv_init} — thermal time from emergence to floral initiation. This parameter determines final leaf number, and thus time to flowering; 2: $t_{flower_to_maturity}$ (°C-d) — thermal time from silking to physiological maturity; 3: $t_{flower_to_start_grain}$ (°C-d) — thermal time from silking to start grain fill period; 4: $potKernelWt$ (g) — potential kernel weight per ear. This parameter sets the maximum kernel weight; 5: $GNmaxCoef$ — coefficient for the tradeoff relationship between grain number and potential kernel weight. This parameter (along with $potKernelWt$) determines maximum number of kernels per ear; 6: rue (g/MJ) — radiation use efficiency; 7: $leaf_app_rate1$ (°C-d) — leaf appearance rate early in the season (typically up to 10th leaf). This parameter influences flowering time; 8: $leaf_app_rate2_3$ (°C-d) — leaf appearance rate later in the season (~10th leaf to flag leaf). This parameter influences flowering time; 9: $nUptakeCease$ (°C-d) — This parameter determines the end of crop N uptake (anywhere from middle to end of grain filling); 10: lai_sla_min — This parameter determines the minimum specific leaf area and has implications for LAI and tissue N % simulations and is an array with two values; 11: $targetLeafSLN$ (g N/m²) — target specific leaf N. This parameter drives canopy photosynthesis. Below the target value, photosynthesis is reduced; 12: $partition_rate_leaf$ — this coefficient determines the dry matter partitioning to leaf; 13: $frac_stem2flower$ — this coefficient determines the dry matter partitioning between stem and rachis (cob); 14: $grainNfillrate$ (mg N grain⁻¹ °C-d⁻¹) — this parameter determines the initial constant rate on grain N uptake (Soufizadeh et al., 2018); 15: $targetRachisNconc$ (%) — this coefficient determines the target N concentration for rachis and has implications for plant N; 16: $targetStemNconc$ (%) — this coefficient determines the target N concentration for stem at APSIM crop stages emergence, floral initiation, flowering, and maturity; 17: $rootdephrate$ (mm/d) — this parameter determines the root front velocity, and affects soil water and nitrogen availability to the plant at each of the 12 APSIM crop stages.

yield, grain number, grain size, harvest index, total above-ground biomass, individual organ biomasses, individual organ nitrogen concentrations and uptakes, leaf number, and leaf area index; we refer to this as calibration set two. In total, 17 APSIM maize model parameters were calibrated to characterize phenological, morphological, and physiological differences among the 15 hybrids. Table 2 lists all the parameter values.

2.8 | Statistical indices for model performance

The goodness of fit was assessed by calculating the root mean square error (RMSE) and normalized root mean square error (NRMSE %; normalization was achieved by dividing the RMSE by the mean of the observed data). These two indices measure the absolute and relative error, respectively. Smaller values of RMSE and NRMSE indicate better model fit. The Index of Agreement (d), a standardized measure of the degree of model prediction error ranging from 0 to 1 with 1 being perfect model fit, was also computed as $1 - \frac{\sum_{i=1}^N (O_i - S_i)^2}{\sum_{i=1}^N (|S_i - \bar{O}| + |O_i - \bar{O}|)^2}$ (Willmott, 1981). The repeatability, R^2 , which is the correlation between repeated measurements was computed as the fraction of variance explained by the model, $1 - \text{Sum}((O_i - S_i)^2) / \text{Sum}((O_i - \bar{O})^2)$ where S_i is the predicted value of O_i , and \bar{O} is the mean of O . Data analysis was completed in R v3.6.1 (R Development Core Team, 2019) using packages APSIM (Fainges, 2019), dplyr (Wickham et al., 2019), and hydroGOF (Zambrano-Bigiarini, 2020).

2.9 | Quantifying data needs for calibration

To better understand data needs and priorities for efficient calibration of hybrids in the model, we performed two levels of reduced calibration: a no calibration control and a partial calibration, which were compared to a full calibration. The no-calibration control was simulated using parameter values from a single commercial hybrid (Pioneer P1197) for all 12 hybrids grown in 2017. Partially calibrated simulations were obtained by utilizing only phenological observations while setting the remaining parameters to values of the commercial hybrid P1197. Full calibration was equivalent to calibration set two (Table 2). By comparing model fits for the three approaches, we quantified the value of phenological, physiological, and N-related data on model performance.

2.10 | Quantifying $G \times E$ on plant traits

To quantify $G \times E$, we used calibration set two to perform long-term (2000–2019) simulations in central Iowa. Historical weather data were sourced from Iowa Environmental Mesonet (Iowa Environmental Mesonet, 2019) using the AGR location (42.0210, -93.7779). Model outputs included end of season grain yield, grain N concentration, harvest index, maximum LAI, and flowering time. Each simulation started on January 1, 2000 and ended December 25, 2019. Soil water, soil N, and surface organic matter was reset on January 1 of each simulated year. Simulated data were analyzed by fitting a mixed linear model in R for each trait using package lme4 (Bates et al., 2015), where environment was fixed, and hybrid was random. The residual in this analysis was assumed to be $G \times E$. Finally, we computed the coefficient of variation (CV = standard deviation/mean) to estimate the range of variability in the simulated plant traits across the 20 years.

3 | RESULTS

3.1 | Observed and simulated phenotypic variance

Across hybrids and environments, observed grain yields at maturity ranged from 2675 to 13,663 kg/ha, maximum above-ground biomass from 14,181 to 26,995 kg/ha, maximum LAI from 3.1 to 5.8 m²/m², and grain N concentration at maturity from 0.79% to 2.04% (Figure 3). The model using calibration set two provided an adequate representation of expected patterns (Figure 3). Simulated grain yields ranged from 8331 to 12,512 kg/ha, maximum above-ground biomass from 17,850 to 25,651 kg/ha, maximum LAI from 4.8 to 6.0 m²/m² and maturity-stage grain N concentration from 0.96% to 1.5% (Figure 3).

3.2 | Model calibration and performance

The developed cultivar parameters using 2017 data only captured the observed variation among hybrids in 2017, with 13 of 16 traits evaluated having an $R^2 > 0.800$ (Table 3). Grain yield, throughout the growing season, had an R^2 value of 0.989 and NRMSE of 16% (Table 3). Total above-ground biomass had an average R^2 of 0.970 and NRMSE of 13% (Table 3). Stalk biomass, senesced leaf biomass, grain size, grain N concentration/uptake, leaf N concentration, and stalk N concentration/uptake had lower repeatability ($R^2 < 0.700$) or higher NRMSE (>25%; Table 3).

TABLE 3 Summary statistical indices of simulated data

Variable	Calibration set 1 2017 environment			Calibration set 1 2018 environments			Calibration set 2 2017 environment			Calibration set 2 all environments		
	R^2	NRMSE	d	R^2	NRMSE	d	R^2	NRMSE	d	R^2	NRMSE	d
Yield (kg/ha)	0.989	15.88	0.99	0.853	73.67	0.93	0.989	15.92	0.99	0.893	48.10	0.97
End of season yield (kg/ha)	0.910	3.742	0.97	0.011	44.96	0.36	0.792	5.97	0.91	0.34	27.04	0.54
Harvest index	0.977	21.12	0.99	0.948	46.75	0.98	0.981	19.09	0.99	0.945	34.81	0.98
Grain number (grains/m ²)	0.983	13.46	0.99	0.935	37.74	0.97	0.986	12.07	0.99	0.924	31.27	0.98
Grain size (g/1000 grains)	0.966	25.70	0.99	0.914	43.86	0.97	0.973	22.96	0.99	0.947	31.62	0.99
Above-ground biomass (g/m ²)	0.970	12.94	0.99	0.902	23.84	0.97	0.969	12.79	0.99	0.929	17.76	0.98
Stalk biomass (g/m ²)	0.858	45.23	0.87	0.805	35.97	0.93	0.809	31.03	0.92	0.764	31.21	0.93
Green leaf biomass (g/m ²)	0.915	17.35	0.97	0.870	17.71	0.94	0.916	14.29	0.98	0.858	19.50	0.95
Senesced leaf biomass (g/m ²)	0.806	57.37	0.93	0.930	47.31	0.93	0.777	60.48	0.92	0.782	56.21	0.92
Leaf area index (m ² /m ²)	0.931	14.99	0.97	0.494	38.78	0.76	0.906	18.40	0.96	0.677	31.11	0.87
Maximum leaf number	0.508	5.59	0.59	0.912	17.05	0.92	0.508	5.59	0.59	0.930	14.86	0.94
Grain N uptake (N g/m ²)	0.892	34.33	0.94	0.673	85.23	0.76	0.897	37.77	0.92	0.651	52.08	0.89
Leaf N uptake (N g/m ²)	0.857	18.14	0.96	0.427	30.65	0.77	0.826	20.23	0.95	0.611	25.73	0.87
Stalk N uptake (N g/m ²)	0.443	68.72	0.65	0.181	90.70	0.41	0.229	49.73	0.63	0.281	48.78	0.69
Leaf N concentration (%)	0.001	43.15	0.47	0.123	34.10	0.14	0.924	11.19	0.97	0.832	14.79	0.95
Grain N concentration (%)	0.197	22.99	0.67	0.051	59.06	0.44	0.294	21.32	0.72	0.298	30.35	0.72
Stalk N concentration (%)	0.906	52.34	0.92	0.933	34.31	0.97	0.966	23.00	0.99	0.950	26.17	0.99
Specific leaf N (N g/m ²)	0.225	19.96	0.59	0.114	20.84	0.52	0.295	20.61	0.66	0.213	22.05	0.61

Notes: y is from $y = ax$ regression; NRMSE: normalized root mean square error (0 to ∞ ; 0 is best; values presented as %; normalization by dividing by mean observed data); d : Index of Agreement— a standardized measure of the degree of model prediction error (0 to 1; 1 is best).

Abbreviation: N, nitrogen.

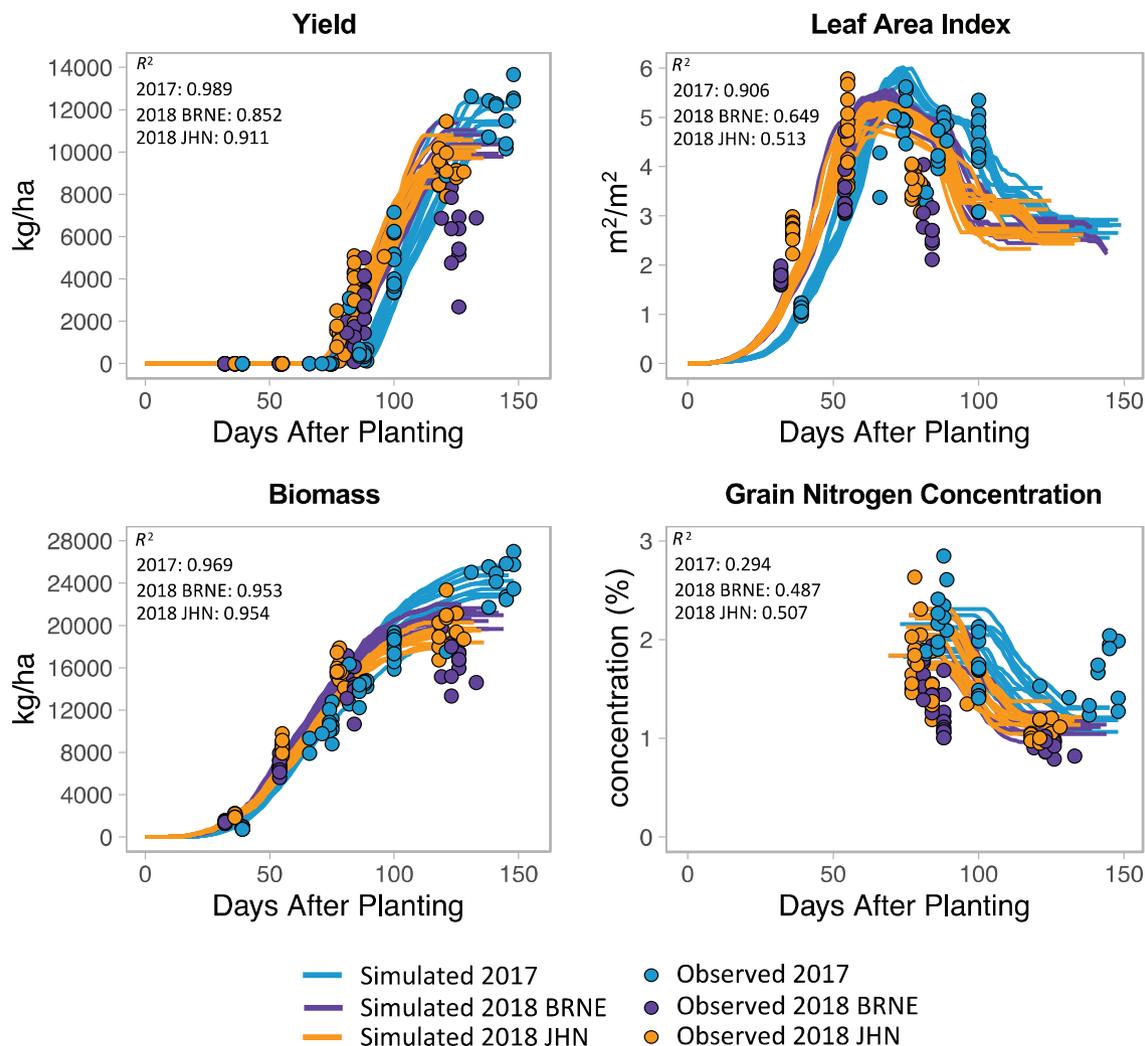


FIGURE 3 Observed (points) and simulated (lines) total above-ground biomass (kg/ha), grain yield (kg/ha), leaf area index (m^2/m^2), and grain N concentration (%) phenotypes across 36 hybrid \times environment combinations, where colors indicate environment. Simulations were performed using calibration set two (all data)

Using the 2017 calibration, we simulated the performance of the nine common hybrids (Table 1) in 2018, a very contrasting year in terms of precipitation (Figure 1). Only three traits (above-ground biomass, maximum leaf number, and green leaf biomass) had an average NRMSE $<25\%$ and an $R^2 > 0.700$ (Table 3). Simulated end of season grain yield was on average 3120 kg/ha above observed grain yield.

Hybrid parameter values were recalibrated using data from all three environments, resulting in calibration set two. The ad hoc modification for waterlogging improved model performance by decreasing NRMSE (%) values for stalk N uptake (-9.21), leaf N concentration (-5.48), stalk N concentration (-3.65), grain N uptake (-3.16), grain yield (-2.82), end of season grain yield (-2.45), senesced leaf biomass (-2.07), green leaf biomass (-1.85), and LAI (-1.42). The model fit results of calibration set two can be found in Table 3, Figure 4, and Figures S2–S36 (individual hybrid \times environment). The most notable improvements to the model after incorporating

all measured data were in the simulation of leaf and stalk N concentration (R^2 increased from 0.001 to 0.924 and 0.906 to 0.966 in 2017, respectively; Table 3). The improvement in leaf dynamics was primarily a result of recalibrating the relationship between the specific leaf area minimum and LAI for all 15 public hybrids (Table 2 lai_sla_min, Figure S37).

Grain yield, when averaged across all hybrids and all environments, had an R^2 value of 0.893 (Table 3). The hybrid with the worst model fit for yield was LH195 \times Mo17 (Figure 5, Figures S18 and S30). This hybrid was grown only in the 2018 environments, which experienced flooding, making it difficult to calibrate given model limitations. However, yield prediction of this hybrid in 2018 JHN environment was good (Figure 5), meaning the lower average model fit for this hybrid ($R^2 = 0.66$) was due primarily to over-simulation in the 2018 BRNE environment (Figure 5). In fact, model predictions of yield were reliably worse across hybrids in the 2018 BRNE environment (NRMSE ranging from 31% to 432%)

B73 x Mo17 Observed vs Simulated 2017

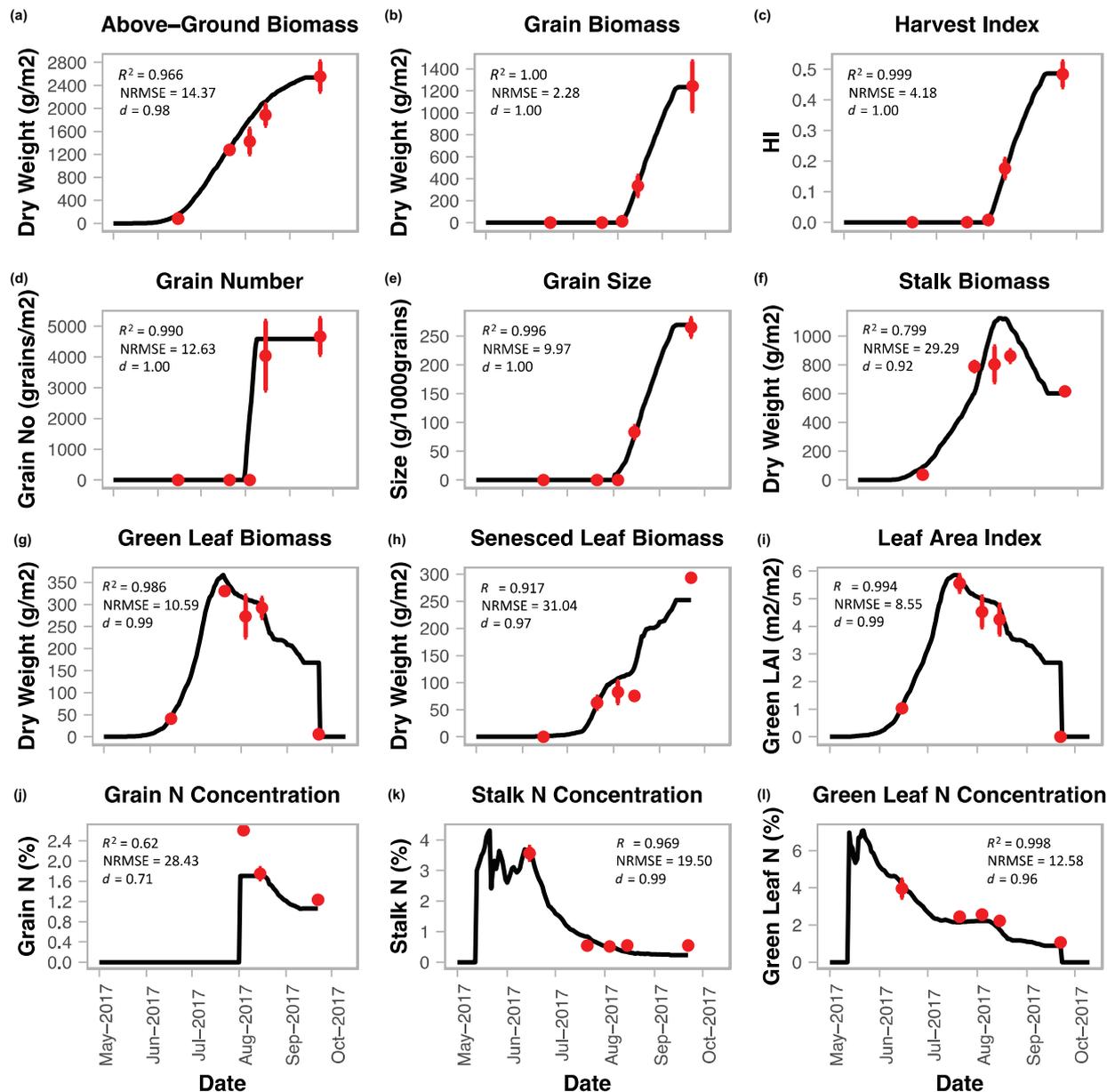


FIGURE 4 Observed versus simulated plots of B73 × Mo17 hybrid in the 2017 environment for 12 traits. Red points represent observed data and corresponding standard errors, while black lines represent simulated values using calibration set two (all data). R^2 , NRMSE (%) and d (Index of Agreement) are reported

compared to the 2017 environment (2%–28%) and 2018 JHN environment (19%–71%), despite higher total rainfall in JHN (Figure 1). Lodging was also most prevalent in 2018 BRNE (Figure 6).

The stalk biomass simulated time series followed observations up to flowering well (Figure 4, Figures S2–S36 panel F). After flowering, the model over-simulated stalk biomass. However, when we compared stalk biomass to observed stalk biomass + shank biomass, the overall R^2

value increased from 0.764 (Table 3) to 0.882, suggesting a code issue in partitioning rules between stem and shank.

Additionally, model fit of grain N concentration was relatively low, evident by an R^2 value of 0.298 when averaged across hybrids and environments (Table 3). In 2017 specifically, observed grain N concentration did not follow a strong dilution curve (Herrmann & Taube, 2004) in six out of the 12 hybrids (B73 × PHZ51, LH195 × PHM49, LH195 × PHZ51,

Observed vs. Simulated Biomass & Yield

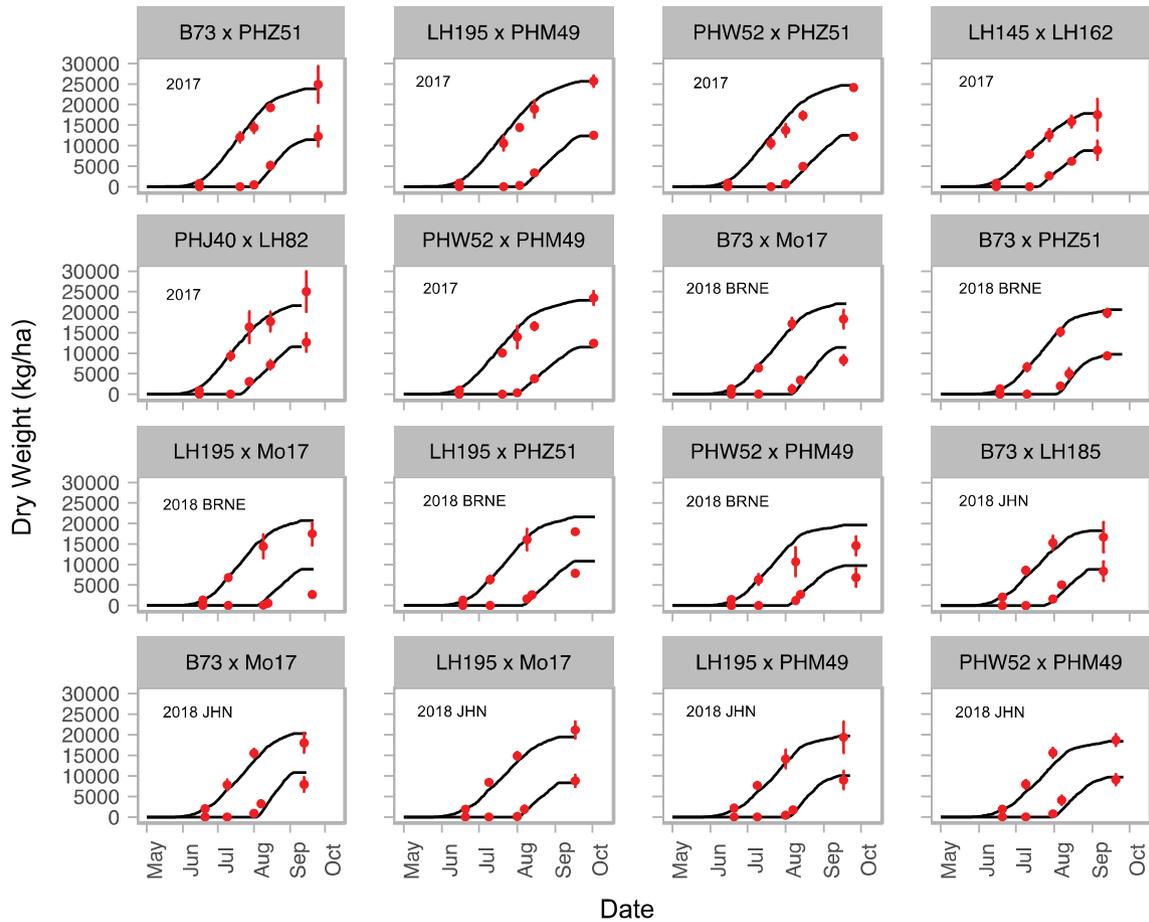


FIGURE 5 Observed versus simulated biomass and yield (kg/ha) for 16 hybrid × environment combinations. Black lines represent simulated values using calibration set two (all data). Red points represent observed data and corresponding standard errors

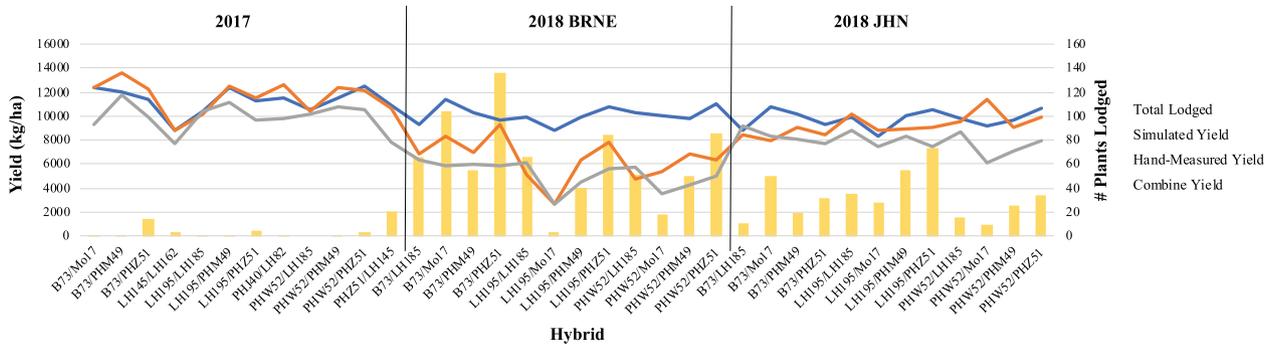


FIGURE 6 Agricultural Production Systems sIMulator (APSIM) simulated yield (calibration set two—all data), hand-harvested yield, combine yield, and number of plants root/stalk lodged in center plot rows for all hybrid × environments. All yields are expressed in kg/ha dry weight. Lodging counts, hand-harvested yields, and combine yields are averaged across replications

PHW52 × LH185, PHW52 × PHZ51, and LH195 × LH185). In other words, at physiological maturity, grain N concentration did not decrease significantly from what it was at the beginning of grain filling, as would be expected. The model did not simulate this pattern (Figures S3–S8 panel J).

3.3 | Variability of model parameters among hybrids

The variability of the calibrated model parameters among hybrids (coefficient of variation) ranged from 0% to 26.8%

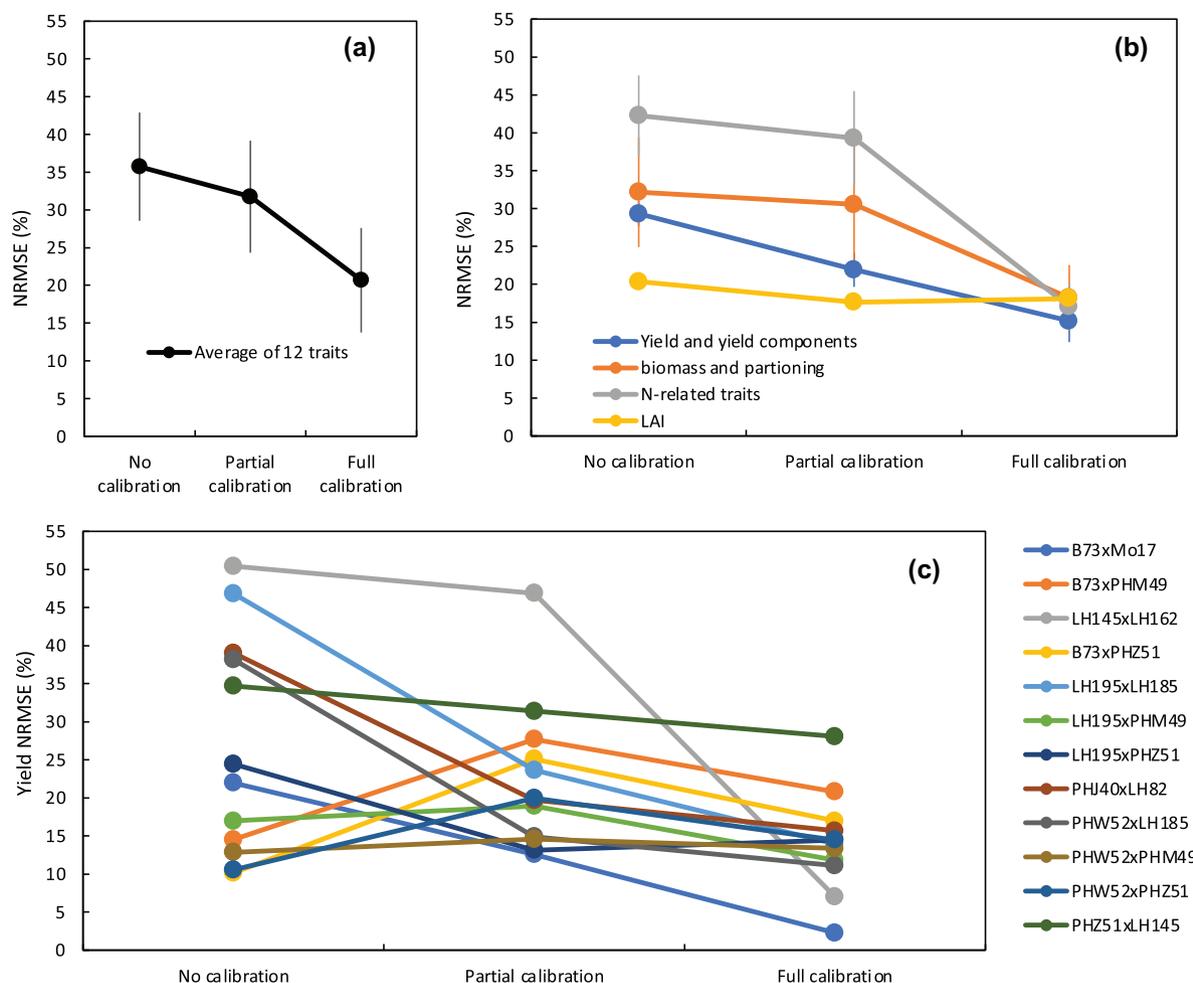


FIGURE 7 Normalized root mean square errors (NRMSE; %) for simulated and observed 2017 data where simulations were run using no calibration, partial calibration, and full calibration. (a) NRMSE values when averaged across 12 traits and 12 hybrids, (b) average NRMSE with traits grouped into four categories, (c) average NRMSE for 12 traits

(Table 2, Figure S1). Phenology parameters, such as thermal time from emergence to end of the juvenile stage (gdd; growing degree days), thermal time from flowering to maturity (gdd), and thermal time from flowering to start of grain fill (gdd) had a coefficient of variation of 6.5%, 8.8%, and 26.8%, respectively (Figure S1). The variation in those parameters is also likely inflated due to three of the hybrids grown in 2017 (LH145 × LH162, PHZ51 × LH145, and PHJ40 × LH82) being of an earlier maturity and different phenology to the other hybrids.

3.4 | Evaluation of data needed for efficient calibration

Compared to no calibration (all parameters from a single commercial hybrid, Pioneer P1197), the partial calibration (phenological parameters only) decreased the NRMSE from 35% to 30%, while the full calibration (all data) decreased the NRMSE to 20% (Figure 7). The calibration process did not have the same impact on all traits. Some traits

improved more than others with calibration (Figure 7). We found that N-related traits benefitted the most from full calibration, followed by biomass and then yield-related traits (Figure 7b). LAI was improved by partial calibration of the crop development parameters but not by full calibration, demonstrating that morphology is driven mostly by environment and phenological parameters and less so by growth parameters (Figure 7b). Individual hybrids responded differently to calibration as well. Taking yield as an example, most hybrids improved with calibration, but a few did not (Figure 7c). Calibration efforts were conducted at a systems-level involving many traits within several environments, so yield predictions were sacrificed at times.

3.5 | Exploring sensitivity of crop parameters

Sensitivity analysis indicated that 49% of the variability in grain yield was due to crop development parameters, 29% due to crop growth parameters, and 22% due to grain component

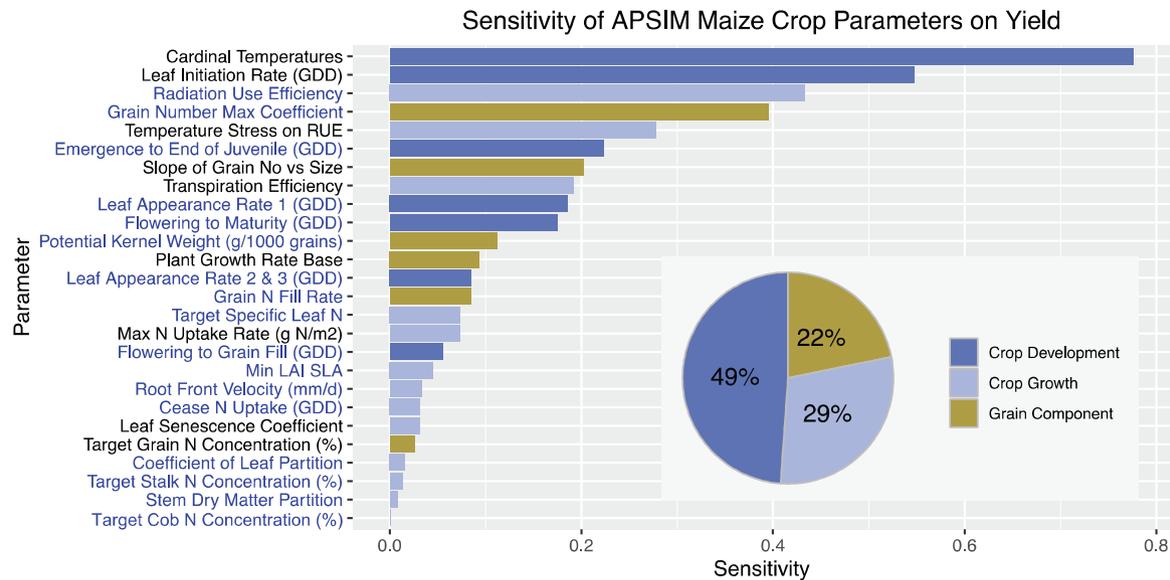


FIGURE 8 Sensitivity of 26 Agricultural Production Systems sIMulator (APSIM) crop cultivar parameters on yield (kg/ha) in terms of relative sensitivity index. Parameters are characterized by crop development, crop growth, and grain component parameters. Parameters that were used in calibration are highlighted in blue text on the y axis. Abbreviation: GDD, growing degree days

parameters (Figure 8). Grain yield variability was most sensitive to cardinal temperatures (minimum, optimum and ceiling temperatures for plant development), with a sensitivity index of 0.78 (Figure 8). RUE was another highly sensitive parameter, with an index of 0.43 (Figure 8). A parameter relating grain number and grain size (grain number max coefficient, Table 2) ranked fourth in sensitivity (sensitivity index = 0.4, Figure 8). Calibration of this parameter improved prediction accuracies of yield, and more specifically, grain size, number, and N concentration (Figure S38). Three of the crop development parameters used in this calibration study were thermal time from emergence to end of juvenile stage, flowering to maturity, and flowering to grain fill. Yield variation is sensitive to these parameters (ranking 6th, 10th, and 17th, respectively), which vary genetically even among the similar maturity range hybrids in this study (Figure 8, Figure S1), reinforcing the need for obtaining phenology estimates in plant breeding data collection.

Other traits such as biomass, crop N uptake, grain number, grain size, grain protein, grain N uptake, harvest index, maximum LAI, and root depth had varying sensitivities to input parameter values (Figures S39–S47). About 54% of the maximum LAI variability was attributed to crop development parameters, with maximum LAI being most sensitive to the thermal time between emergence and end of juvenile stage. Harvest index was most sensitive to crop development parameters (61%), followed by grain component parameters (24%) and crop growth parameters (15%). Of the 10 output traits evaluated in the sensitivity analyses, crop development parameters were, on average, attributed to most of the trait

sensitivity, ranging from 26% to 61% (Figure 8, Figures S39–S44).

3.6 | Quantifying G×E

Simulation analysis for 20 weather years showed strong crossover $G \times E$ among the 15 hybrids (Figure 9a). $G \times E$ accounted for 19% of the total genotypic variance for end of season grain yield, 4% for maximum LAI, 17% for grain N concentration, 27% for harvest index, and 2% for flowering time. B73 × Mo17 had the highest average simulated end of season grain yield (11,661 kg/ha) across the 20 environments but ranked eighth in variation among the 15 hybrids, with a coefficient of variation of 6.1% (Table S4). This variation indicates that while B73 × Mo17 may perform better than other hybrids on average, it may not always perform the highest, suggesting a potential weakness under certain environmental conditions. Across hybrids, the coefficient of variation ranged from 3.5% to 10.1% for end of season grain yield, 10.7% to 15.3% for grain N concentration, 4.1% to 8.5% for harvest index, 1.9% to 3.9% for maximum LAI, and 5.6% to 6.1% for flowering days after planting.

4 | DISCUSSION

4.1 | Modeling physiological variability

Our results demonstrate that calibrated CGMs can capture a portion of the observed variability among hybrids, but

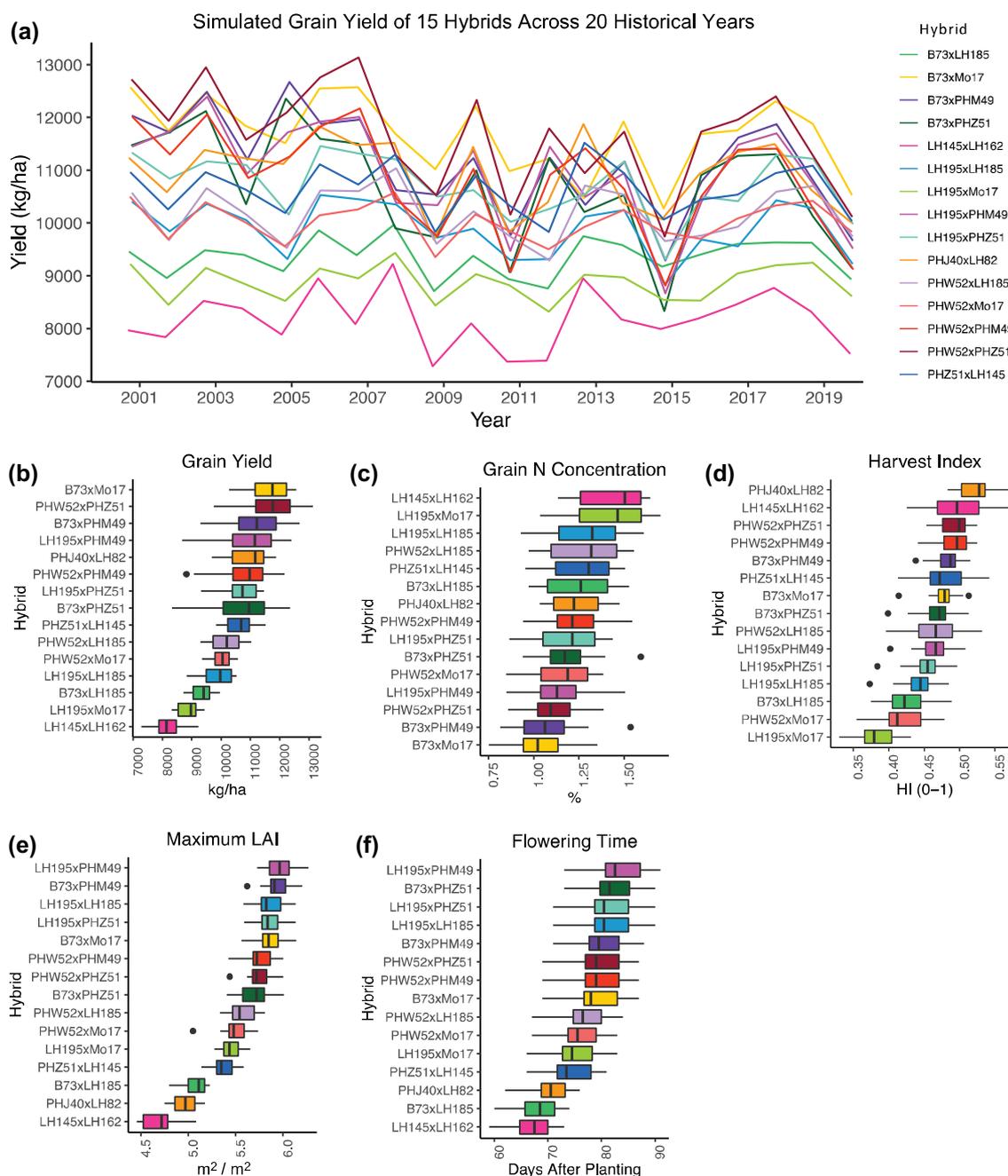


FIGURE 9 Simulations of end of season grain yield (a and b), grain N concentration (c), harvest index (d), maximum LAI (e), flowering time (f) for all hybrids using calibration set two and historical weather data from 2000 – 2019 in Ames, IA, environment

not all the variability (Figure 3). While more physiological, detailed plant processes have been incorporated within CGMs (Hammer et al., 2010), it is a common notion that more is needed (Wang et al., 2019), especially with regard to lodging and excessive moisture stress conditions in which we found the largest model inaccuracies in this study (Figure 6). The relatively small observed phenotypic variability among the hybrids in this study (Figure 3) reflects the limited genetic diversity among breeding population material and hybrids grown in the U.S. Corn Belt today. Many commercial hybrids

are derived from the same public inbred parents, such as the ex-PVP inbreds used in this study (Coffman et al., 2020; Darrah & Zuber, 1986; Mikel & Dudley, 2006). Therefore, CGMs need to simulate subtle physiological differences among cultivars to capture performance variation among commercial maize hybrids. A balance of parsimony is necessary, though, if CGMs are to play an integral part in plant breeding (Hammer et al., 2019; Messina et al., 2018) and linked directly to genes or quantitative trait loci (QTLs; Hammer et al., 2016; Yin et al., 2016). Until an understanding of what

underlying physiological traits differ across genotypes for hybrid performance or $G \times E$ is reached, the amount of variation that can be predicted cannot truly be determined.

In this study, full calibration resulted in R^2 values greater than 0.75 for 11 out of 18 traits evaluated across all hybrids and environments (Table 3). However, the overall R^2 and NRMSE (%) results suggest that we could not capture the N content patterns well in this study (Table 3). It is important to note, though, that the model fit results related to N may be biased from having few data at early stages of development for leaf, stalk, and grain N content (Figure 4) and because the values and variation in N content are small. Furthermore, error values for ratios (organ N uptake) are perpetually greater than that of their constituent parts (organ biomass and N concentration; Table 3). Although R^2 is a statistic widely used for model evaluation, it is sensitive to outliers and insensitive to additive and proportional differences among observed and predicted data (Archontoulis et al., 2014; Legates & McCabe, 1999). Therefore, we include the Index of Agreement (d) statistic which can detect additive and proportional differences (Table 3) (Legates & McCabe, 1999; Willmott, 1981).

4.2 | Data requirement learnings for hybrid calibration

As demonstrated in this study, phenology is important for simulating hybrid performance (reduced average NRMSE from 35% to 30%, Figure 7), while crop growth and grain component parameters account for a portion of the physiological variability among hybrids (further reduced average NRMSE to 20%, Figure 7). These results support the development of generic cultivars based primarily on phenological observations as done in the past (Boote et al., 2003) but demonstrate the benefit of additional trait data collection and calibration. Sensitivity analyses revealed important parameters for trait variability and future consideration for data collection in plant breeding programs and phenotyping experiments (Figure 8). The sensitivity analyses in this study were conducted using simulations of a single environment, which may limit the repeatability of results across environments. Additional work is planned to address the environmental sensitivity of the model parameters. Sensitivity analyses can reveal strengths and weaknesses in both the calibration protocol, that is, the data, and in the CGM. Phenotypic sensitivity to particular parameters may reveal that those parameters are critical for accurate simulations or that the structure of the model does not accurately represent the effect of the parameter.

Calibration studies and sensitivity analyses provide a better understanding of the phenotypic data requirements for model calibrations, insight into target CGM parameters for calibration in an optimization schema (Wallach et al., 2021),

and identification of important physiological traits underlying hybrid performance that could be the focus of direct phenotyping experiments performed on a large number of entries (Cooper, Messina, et al., 2014). For instance, the high yield sensitivity of the RUE parameter (Figure 8) suggests that accurate phenotyping of this trait would significantly improve simulation, and there is evidence RUE has increased through maize breeding efforts (Curin et al., 2020; Tollenaar & Aguilera, 1992). Our results also demonstrate the importance of crop development in the simulation of plant traits, suggesting that phenology data collection and proper calibration of phenology parameters can account for nearly 49% of the yield variability (Figure 8). Phenology data, even at the scale of large plant breeding trials, is not difficult to obtain. Specifically, thermal time from emergence to end of juvenile stage, thermal time from flowering to maturity, and leaf initiation and appearance rates were demonstrated to have high yield and above-ground biomass sensitivity (Figure 8, Figure S55). Measuring end of season grain N concentration could be beneficial to model calibration as well, as our results suggest grain N concentration parameters may be hybrid specific given the observed dilution curves (Figures S3–S8 panel J), rather than crop specific as defined in the APSIM model. Grain N concentration is known to vary among hybrids (DeBruin et al., 2017; Yan et al., 2014; Zhang et al., 2020) and is a trait easily obtained even at the scale of plant breeding programs, especially once combine harvesters are improved to provide both grain yield and protein (Long & McCallum, 2020). Our results thereby suggest that more work is needed in the area of N content through a combined approach of data collection and model improvement.

The time and labor associated with obtaining phenotypic data needed for full empirical calibration is a significant hurdle. However, as advancements in high-throughput phenomics are made, we anticipate the ability to estimate additional model parameters more efficiently using imaging and other high-throughput phenomic technologies (Kusmec et al., 2021). Phenomics and CGMs will continue to undergo joint development as phenotypic data are used to improve the models and CGMs guide phenotyping efforts through physiological understanding and identification of knowledge gaps (Hammer et al., 2002). Advances in phenomics in concert with methods for imputation of CGM parameters will enable further advancement in calibration protocols (Messina et al., 2018).

4.3 | Model structure limitations—Waterlogging

Model structure limitations relating to waterlogging were encountered in this study in the 2018 environments (Figures 1 and 6). Waterlogging, characterized by the soil water being

near saturation, has been shown to decrease dry matter accumulation, leaf photosynthesis, LAI, plant height, stalk strength, and root growth, while increasing leaf senescence (Araki et al., 2012; Hayashi et al., 2013; Kanwar et al., 1988; Ren et al., 2016b, 2016c; Ren et al., 2014; Tian et al., 2020). These responses can cause lodging, root rot, fungal disease infections, and overall grain yield reductions (Kettlewell et al., 1999). While waterlogging functions have been tested within the APSIM soybean model version 7.9 (Pasley et al., 2020), APSIM maize version 7.10 currently accounts for only root depth inhibition due to excess moisture (Ebrahimi-Mollabashi et al., 2019). Thus, we performed an ad hoc modification to partially account for excess water moisture in this study (see Section 2). Although the model modification helped, it did not solve the yield over-prediction issue in year 2018 where waterlogging was prevalent in early vegetative stages when maize is most susceptible (Mukhtar et al., 1990; Ren et al., 2016a; Zaidi et al., 2004) and late-season stalk and root lodging were noted (Figure 6). More work is needed to expand modeling capacity in wet soils. The need for such model enhancement is further justified by historic crop insurance payments in the U.S. Corn Belt, which reveal 30% of crop damage due to excess water (Perry et al., 2020). Globally, 10%–12% of agricultural land is estimated to be affected by waterlogging (Kaur et al., 2019). These numbers are predicted to increase because of frequent, severe, and unpredictable weather events associated with climate change (Hirabayashi et al., 2013; IPCC, 2014), which reinforces the need to improve model structure to deal with excess water stress.

4.4 | Simulating G×E across 20 years

A calibrated CGM that simulates yield among hybrids, providing continuous data rather than a few single time point measurements (Figure 3), holds immense potential for plant breeders to understand cultivar differences and G × E. We demonstrated that a CGM calibrated using empirical data simulates the kind of G × E expected and observed by breeders, a result that helps validate the CGM structure (Figure 9). The hybrid B73 × Mo17 performed better in terms of grain yield than the other hybrids averaged across 20 years but was more sensitive to environmental variation than the other hybrids. G × E accounted for 2%–29% of the total genotypic variance across the five traits evaluated, and for yield, G × E accounted for 19%. These numbers are consistent with G × E proportions of variance observed in multi-environment trials of commercial hybrids grown in Iowa (So & Edwards, 2009), synthetic maize population crosses in Iowa (Edwards, 2016), and hybrids grown in G2F network (Gage et al., 2017). If CGM calibrations are validated with multi-environment trial data, the ability to simulate differences in environmental stability would be a transformational tool in plant breeding.

It would enable breeders to use calibrations developed in a very limited set of environments, including within a single year, and potentially extrapolate to many environments and years to identify stable hybrids. This level of environmental extrapolation is currently not possible with linear statistical prediction models. Multi-environment validation of these CGM calibrations is currently underway.

4.5 | Potential limitations and future considerations

Within the modeling community, consensus on the optimal approach to model calibration has not been obtained because no single method is best in every context (Seidel et al., 2018; Wallach et al., 2021). Even for the same model structure and data set, different modeling groups may identify different sets of parameters to estimate, different parameter values, and different criteria for defining the best model parameters (Wallach et al., 2021). We used manual calibration, expert knowledge and sensitivity analyses to calibrate models with a combination of visual and frequentist best-fit metrics to determine final parameter values. We recognize both the merits and limitations to our approach. We used a stepwise approach, calibrating parameters in order of importance, starting first with phenology, then biomass production, biomass partitioning, then tissue N, and repeated the process to ensure the model produced sensible results (Figure 2). Our approach was similar to that of an optimizer in that it was iterative in nature as parameter values were adjusted based on model predictions for multiple traits. Whereas a manual process does not satisfy formal optimality criteria, it does avoid some potential pitfalls in automated approaches (Wallach et al., 2021). Limitations of automated approaches can include lack of directly-linked optimizer software for certain modeling engines like APSIM classic, convergence to local rather than global optima, and choice of initial model conditions impacting reproducibility. Even with automated methods, calibration requires knowledge of the model structure and requires sensible assumptions for initial conditions, and hence, some level of manual input is still required. Equifinality is a common problem with the calibration of many parameters and occurs when different combinations of parameter values give the same results; therefore, the calibration does not offer unique parameter values (He et al., 2017; Wallach et al., 2021). While the calibrations in this study were based on a limited number of environments, there was contrast among the environments in relation to water stress in particular, which could reduce some equifinality impact and boost robustness (He et al., 2017). Future research will have the potential to improve these calibrations based on future model improvements, improvements to calibration procedures, and additional data on the same hybrids (some of which is currently being collected in our lab), and

thus these calibrations are good starting points for a set of publicly available U.S. Corn Belt hybrids.

CGM calibration and simulation studies can help to identify important underlying traits and provide physiological insight for future applications within breeding. Applications of CGMs within breeding may be more plausible at certain stages within the breeding pipeline where fewer candidates need to be evaluated. In late stages of breeding programs, CGMs could be utilized to characterize important parent or tester lines, for product placement, or to identify important traits within managed stress nurseries (Cooper, Messina, et al., 2014).

5 | CONCLUSIONS

The sensitivity analyses and observations in this study elucidate the importance of phenology underlying differences among hybrids and the need for this information to be measured in breeding experiments. Particularly phenology traits such as leaf initiation and appearance rates, the thermal time from emergence to end of juvenile stage, and the thermal time from flowering to maturity were shown to be important. Yield and above-ground biomass are also largely sensitive to RUE, which suggests that the overall simulation accuracy could be improved if RUE measurements were collected. By calibrating 15 public hybrids within the APSIM model and providing information on data needs for efficient calibration and importance of traits, this study creates future opportunities for increased model use in breeding programs. We also recognize, however, that CGMs are likely to be incorporated in breeding programs by imputing genetic parameters through optimization schema due to the reduced need for extensive phenotypic data (Cooper, Messina, et al., 2014; Messina et al., 2018). The development of parameter values particularly for public hybrids, as done in this study where genetic marker and pedigree data are available, provides opportunities for future research to link plant traits to genes. These calibrations provide a range of possible values for these historical hybrids which can be used in future exploration studies by APSIM users. Specifically, given that these hybrids have been grown in many environments throughout the United States as part of the G2F Initiative (McFarland et al., 2020), it creates opportunities to further test model simulations across a range of environments, use the model to assist empirical research by proving insights on water-nitrogen limitations per environment, simulate G × E in the long term (as we demonstrated in this study), or even explore the role of management on hybrid performance to better understand agronomic yield gaps.

ACKNOWLEDGMENTS

This research was supported in part by the US. Department of Agriculture, Agricultural Research Service (5030-21000-

066-00D), AFRI (2017-67013-26463), USDA Hatch project (IOW04614, IOW10480), NSF (DGE-1545453, #1830478, #1842097), FFAR (#534264, DSnew-0000000014) and Iowa State University Department of Agronomy. The findings and conclusions in this publication are those of the authors and should not be construed to represent any official USDA or U.S. Government determination or policy. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture. USDA is an equal opportunity provider and Employer. We thank Iowa State University undergraduate students (Hannah Hegwer, Dirk Winkelman, Brian Today, Aubry Grimm, Miranda Mathiason, Ellen Reed, Kayla Morgan, Ethan Ley), post-doctoral Dr. Raziél A. Ordóñez, and research technician John Golden for assistance with data collection. We also thank Iowa State University Agricultural Engineering and Agronomy Research Farm managers and technicians for assistance in managing field experiments. Acknowledgment is also made to the APSIM Initiative for making the software publicly available and ensuring software quality.

AUTHOR CONTRIBUTIONS

Data curation, formal analysis, investigation, methodology, writing-original draft, and writing-review and editing: Cassandra Anne Winn. Conceptualization, formal analysis, funding acquisition, methodology, resources, and writing-review and editing: Sotirios Archontoulis. Conceptualization, data curation, funding acquisition, methodology, project administration, resources, supervision, and writing-review and editing: Jode Edwards.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ORCID

Sotirios Archontoulis  <https://orcid.org/0000-0001-7595-8107>

Jode Edwards  <https://orcid.org/0000-0001-7918-476X>

REFERENCES

- Abendroth, L. J., Elmore, R. W., Boyer, M. J., & Marlay, S. K. (2011). *Corn growth and development* (Vol. PMR 1009). Iowa State University.
- Anderson, II, S. L., Murray, S. C., Malambo, L., Ratcliff, C., Popescu, S., Cope, D., & Thomasson, J. A. (2019). Prediction of maize grain yield before maturity using improved temporal height estimates of unmanned aerial systems. *The Plant Phenome Journal*, 2(1), 190004. <https://doi.org/10.2135/tppj2019.02.0004>
- Araki, H., Hamada, A., Hossain, M. A., & Takahashi, T. (2012). Water-logging at jointing and/or after anthesis in wheat induces early leaf senescence and impairs grain filling. *Field Crops Research*, 137, 27–36. <https://doi.org/10.1016/j.fcr.2012.09.006>

- Archontoulis, S. V., Castellano, M. J., Licht, M. A., Nichols, V., Baum, M., Huber, I., & Lamkey, K. R. (2020). Predicting crop yields and soil-plant nitrogen dynamics in the US Corn Belt. *Crop Science*, 60(2), 721–738. <https://doi.org/10.1002/csc2.20039>
- Archontoulis, S. V., Miguez, F. E., & Moore, K. J. (2014). Evaluating APSIM maize, soil water, soil nitrogen, manure, and soil temperature modules in the Midwestern United States. *Agronomy Journal*, 106(3), 1025–1040. <https://doi.org/10.2134/agnonj2013.0421>
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Scientific Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Baum, M. E., Licht, M. A., Huber, I., & Archontoulis, S. V. (2020). Impacts of climate change on the optimum planting date of different maize cultivars in the central US Corn Belt. *European Journal of Agronomy*, 119, 126101. <https://doi.org/10.1016/j.eja.2020.126101>
- Beckett, T. J., Morales, A. J., Koehler, K. L., & Rocheford, T. R. (2017). Genetic relatedness of previously plant-variety-protected commercial maize inbreds. *PLoS One*, 12(12), e0189277. <https://doi.org/10.1371/journal.pone.0189277>
- Birch, C. J., Rickert, K. G., & Hammer, G. L. (1998). Modelling leaf production and crop development in maize (*Zea mays* L.) after tassel initiation under diverse conditions of temperature and photoperiod. *Field Crops Research*, 58(2), 81–95. [https://doi.org/10.1016/S0378-4290\(98\)00087-2](https://doi.org/10.1016/S0378-4290(98)00087-2)
- Boote, K. J., Jones, J. W., Batchelor, W. D., Nafziger, E. D., & Myers, O. (2003). Genetic coefficients in the CROPGRO-soybean model: Links to field performance and genomics. *Agronomy Journal*, 95(1), 32–51. <https://doi.org/10.2134/agnonj2003.0032>
- Burgueno, J., de los Campos, G., Weigel, K., & Crossa, J. (2012). Genomic prediction of breeding values when modeling genotype x environment interaction using pedigree and dense molecular markers. *Crop Science*, 52(2), 707–719. <https://doi.org/10.2135/cropsci2011.06.0299>
- Bustos-Korts, D., Boer, M. P., Malosetti, M., Chapman, S., Chenu, K., Zheng, B. Y., & van Eeuwijk, F. (2019). Combining crop growth modeling and statistical genetic modeling to evaluate phenotyping strategies. *Frontiers in Plant Science*, 10, 1491. <https://doi.org/10.3389/fpls.2019.01491>
- Bustos-Korts, D., Malosetti, M., Chenu, K., Chapman, S., Boer, M. P., Zheng, B. Y., & van Eeuwijk, F. A. (2019). From QTLs to adaptation landscapes: Using genotype-to-phenotype models to characterize G x E over time. *Frontiers in Plant Science*, 10, 1540. <https://doi.org/10.3389/fpls.2019.01540>
- Carberry, P. S., Muchow, R. C., & Mccown, R. L. (1989). Testing the CERES-Maize simulation-model in a semi-arid tropical environment. *Field Crops Research*, 20(4), 297–315. [https://doi.org/10.1016/0378-4290\(89\)90072-5](https://doi.org/10.1016/0378-4290(89)90072-5)
- Chapman, S., Cooper, M., Hammer, G., & Butler, D. G. (2000). Genotype by environment interactions affecting grain sorghum. II. Frequencies of different seasonal patterns of drought stress are related to location effects on hybrid yields. *Australian Journal of Agricultural Research*, 51(2), 209–221. <https://doi.org/10.1071/Ar99021>
- Chapman, S., Cooper, M., Podlich, D., & Hammer, G. (2003). Evaluating plant breeding strategies by simulating gene action and dryland environment effects. *Agronomy Journal*, 95(1), 99–113. <https://doi.org/10.2134/agnonj2003.0099>
- Chenu, K., Porter, J. R., Martre, P., Basso, B., Chapman, S. C., Ewert, F., & Asseng, S. (2017). Contribution of crop models to adaptation in wheat. *Trends in Plant Science*, 22(6), 472–490. <https://doi.org/10.1016/j.tplants.2017.02.003>
- Coffman, S. M., Hufford, M. B., Andorf, C. M., & Lubberstedt, T. (2020). Haplotype structure in commercial maize breeding programs in relation to key founder lines. *Theoretical and Applied Genetics*, 133(2), 547–561. <https://doi.org/10.1007/s00122-019-03486-y>
- Cooper, M., Gho, C., Leafgren, R., Tang, T., & Messina, C. (2014). Breeding drought-tolerant maize hybrids for the US Corn-Belt: Discovery to product. *Journal of Experimental Botany*, 65(21), 6191–6204. <https://doi.org/10.1093/jxb/eru064>
- Cooper, M., Messina, C. D., Podlich, D., Totir, L. R., Baumgarten, A., Hausmann, N. J., & Graham, G. (2014). Predicting the future of plant breeding: Complementing empirical evaluation with genetic prediction. *Crop & Pasture Science*, 65(4), 311–336. <https://doi.org/10.1071/Cp14007>
- Cooper, M., Technow, F., Messina, C., Gho, C., & Totir, L. R. (2016). Use of crop growth models with whole-genome prediction: Application to a maize multi-environment trial. *Crop Science*, 56(5), 2141–2156. <https://doi.org/10.2135/cropsci2015.08.0512>
- Cooper, M., van Eeuwijk, F. A., Hammer, G. L., Podlich, D. W., & Messina, C. (2009). Modeling QTL for complex traits: Detection and context for plant breeding. *Current Opinion in Plant Biology*, 12(2), 231–240. <https://doi.org/10.1016/j.pbi.2009.01.006>
- Crossa, J., de los Campos, G., Maccaferri, M., Tuberosa, R., Burgueño, J., & Pérez-Rodríguez, P. (2016). Extending the marker x environment interaction model for genomic-enabled prediction and genome-wide association analysis in durum wheat. *Crop Science*, 56(5), 2193–2209. <https://doi.org/10.2135/cropsci2015.04.0260>
- Curin, F., Severini, A. D., Gonzalez, F. G., & Otegui, M. E. (2020). Water and radiation use efficiencies in maize: Breeding effects on single-cross Argentine hybrids released between 1980 and 2012. *Field Crops Research*, 246, 107683. <https://doi.org/10.1016/j.fcr.2019.107683>
- Darrah, L. L., & Zuber, M. S. (1986). 1985 United-States farm maize germplasm base and commercial breeding strategies. *Crop Science*, 26(6), 1109–1113. <https://doi.org/10.2135/cropsci1986.0011183x002600060004x>
- DeBruin, J. L., Schussler, J. R., Mo, H., & Cooper, M. (2017). Grain yield and nitrogen accumulation in maize hybrids released during 1934 to 2013 in the US Midwest. *Crop Science*, 57(3), 1431–1446. <https://doi.org/10.2135/cropsci2016.08.0704>
- Dietzel, R., Liebman, M., Ewing, R., Helmers, M., Horton, R., Jarchow, M., & Archontoulis, S. (2016). How efficiently do corn- and soybean-based cropping systems use water? A systems modeling analysis. *Global Change Biology*, 22(2), 666–681. <https://doi.org/10.1111/gcb.13101>
- Dingkuhn, M., Devries, F. W. T. P., & Miezán, K. M. (1993). Improvement of rice plant type concepts: Systems research enables interaction of physiology and breeding. In *Systems approaches for agricultural development* (vol. 2, pp. 19–35). Springer. https://doi.org/10.1007/978-94-011-2840-7_2
- Donald, C. M. (1968). Breeding of crop ideotypes. *Euphytica*, 17(3), 385–403. <https://doi.org/10.1007/Bf00056241>
- Ebrahimi-Mollabashi, E., Huth, N. I., Holzwoth, D. P., Ordonez, R. A., Hatfield, J. L., Huber, I., & Archontoulis, S. V. (2019). Enhancing APSIM to simulate excessive moisture effects on root growth. *Field Crops Research*, 236, 58–67. <https://doi.org/10.1016/j.fcr.2019.03.014>
- Edmeades, G. O., & Daynard, T. B. (1979). Relationship between final yield and photosynthesis at flowering in individual maize plants.

- Canadian Journal of Plant Science*, 59(3), 585–601. <https://doi.org/10.4141/cjps79-097>
- Edwards, J. W. (2016). Genotype x environment interaction for plant density response in maize (*Zea mays* L.). *Crop Science*, 56(4), 1493–1505. <https://doi.org/10.2135/cropsci2015.07.0408>
- Fainges, J. (2019). APSIM: General utility functions for the ‘agricultural production systems simulator’ R package (Version 0.9.3). <https://CRAN.R-project.org/package=APSIM>
- Furbank, R. T., & Tester, M. (2011). Phenomics—Technologies to relieve the phenotyping bottleneck. *Trends in Plant Science*, 16(12), 635–644. <https://doi.org/10.1016/j.tplants.2011.09.005>
- Gage, J. L., Jarquin, D., Romay, C., Lorenz, A., Buckler, E. S., Kaeppeler, S., & de Leon, N. (2017). The effect of artificial selection on phenotypic plasticity in maize. *Nature Communications*, 8, 1348. <https://doi.org/10.1038/s41467-017-01450-2>
- Godfray, H. C. J., Beddington, J. R., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F., Pretty, J., Robinson, S., Thomas, S. M., & Toulmin, C. (2010). Food security: The challenge of feeding 9 billion people. *Science*, 327, 1–23. <https://www.science.org/doi/10.1126/science.1185383>
- Hamby, D. M. (1994). A review of techniques for parameter sensitivity analysis of environmental-models. *Environmental Monitoring and Assessment*, 32(2), 135–154. <https://doi.org/10.1007/Bf00547132>
- Hammer, G. (2020). The roles of credibility and transdisciplinarity in modelling to support future crop improvement. *in silico Plants*, 2(1), diaa004. <https://doi.org/10.1093/insilicoplants/diaa004>
- Hammer, G., Cooper, M., Tardieu, F., Welch, S., Walsh, B., van Eeuwijk, F., & Podlich, D. (2006). Models for navigating biological complexity in breeding improved crop plants. *Trends in Plant Science*, 11(12), 587–593. <https://doi.org/10.1016/j.tplants.2006.10.006>
- Hammer, G., Kropff, M. J., Sinclair, T. R., & Porter, J. R. (2002). Future contributions of crop modelling—From heuristics and supporting decision making to understanding genetic regulation and aiding crop improvement. *European Journal of Agronomy*, 18(1–2), 15–31. [https://doi.org/10.1016/S1161-0301\(02\)00093-X](https://doi.org/10.1016/S1161-0301(02)00093-X)
- Hammer, G., McLean, G., Chapman, S., Zheng, B. Y., Doherty, A., Harrison, M. T., & Jordan, D. (2014). Crop design for specific adaptation in variable dryland production environments. *Crop & Pasture Science*, 65(7), 614–626. <https://doi.org/10.1071/Cp14088>
- Hammer, G., Messina, C., van Oosterom, E., Chapman, S., Singh, V., Borrell, A., & Cooper, M. (2016). Molecular breeding for complex adaptive traits: How integrating crop ecophysiology and modelling can enhance efficiency. In X. Yin & P. C. Struik (Eds.), *Crop systems biology: Narrowing the gaps between crop modelling and genetics* (pp. 147–162). Springer International Publishing. https://doi.org/10.1007/978-3-319-20562-5_7
- Hammer, G., Messina, C., Wu, A., & Cooper, M. (2019). Biological reality and parsimony in crop models—Why we need both in crop improvement! *in silico Plants*, 1(1), diz010. <https://doi.org/10.1093/insilicoplants/diz010>
- Hammer, G., van Oosterom, E., McLean, G., Chapman, S. C., Broad, I., Harland, P., & Muchow, R. C. (2010). Adapting APSIM to model the physiology and genetics of complex adaptive traits in field crops. *Journal of Experimental Botany*, 61(8), 2185–2202. <https://doi.org/10.1093/jxb/erq095>
- Haverkort, A. J., & Kooman, P. L. (1997). The use of systems analysis and modelling of growth and development in potato ideotyping under conditions affecting yields. *Euphytica*, 94(2), 191–200. <https://doi.org/10.1023/A:1002965428704>
- Hayashi, T., Yoshida, T., Fujii, K., Mitsuya, S., Tsuji, T., Okada, Y., & Yamauchi, A. (2013). Maintained root length density contributes to the waterlogging tolerance in common wheat (*Triticum aestivum* L.). *Field Crops Research*, 152, 27–35. <https://doi.org/10.1016/j.fcr.2013.03.020>
- He, D., Wang, E., Wang, J., & Robertson, M. J. (2017). Data requirement for effective calibration of process-based crop models. *Agricultural and Forest Meteorology*, 234–235, 136–148. <https://doi.org/10.1016/j.agrformet.2016.12.015>
- Hebert, Y. (1990). Genetic-variation of the rate of leaf appearance in maize—Possible yield prediction at the early stage. *Euphytica*, 46(3), 237–247. <https://doi.org/10.1007/Bf00027223>
- Herrmann, A., & Taube, F. (2004). The range of the critical nitrogen dilution curve for maize (*Zea mays* L.) can be extended until silage maturity. *Agronomy Journal*, 96(4), 1131–1138. <https://doi.org/10.2134/agronj2004.1131>
- Heslot, N., Akdemir, D., Sorrells, M. E., & Jannink, J. L. (2014). Integrating environmental covariates and crop modeling into the genomic selection framework to predict genotype by environment interactions. *Theoretical and Applied Genetics*, 127(2), 463–480. <https://doi.org/10.1007/s00122-013-2231-5>
- Hirabayashi, Y., Mahendran, R., Koiraal, S., Konoshima, L., Yamazaki, D., Watanabe, S., & Kanae, S. (2013). Global flood risk under climate change. *Nature Climate Change*, 3(9), 816–821. <https://doi.org/10.1038/Nclimate1911>
- Holzworth, D. P., Huth, N. I., Devoil, P. G., Zurcher, E. J., Herrmann, N. I., McLean, G., & Keating, B. A. (2014). Apsim—Evolution towards a new generation of agricultural systems simulation. *Environmental Modelling & Software*, 62, 327–350. <https://doi.org/10.1016/j.envsoft.2014.07.009>
- Hunter, J. L., Tekrony, D. M., Miles, D. F., & Egli, D. B. (1991). Corn seed maturity indicators and their relationship to uptake of c-14 assimilate. *Crop Science*, 31(5), 1309–1313. <https://doi.org/10.2135/cropsci1991.0011183x003100050045x>
- Huth, N. I., Bristow, K. L., & Verburg, K. (2012). Swim3: Model use, calibration, and validation. *Transactions of the ASABE*, 55(4), 1303–1313. <https://doi.org/10.13031/2013.42243>
- Iowa environmental mesonet. (2019). *Iowa Ag climate network*. Ames, IA: Iowa State University. <https://mesonet.agron.iastate.edu/agclimate/>
- IPCC. (2014). *Climate change 2014: synthesis report. Contribution of working groups i, ii and iii to the fifth assessment report of the intergovernmental panel on climate change*. Geneva, Switzerland: IPCC. <https://doi.org/10.1017/CBO9781107415416>
- Jackson, P., Robertson, M., Cooper, M., & Hammer, G. (1996). The role of physiological understanding in plant breeding; from a breeding perspective. *Field Crops Research*, 49(1), 11–37. [https://doi.org/10.1016/S0378-4290\(96\)01012-X](https://doi.org/10.1016/S0378-4290(96)01012-X)
- Jarquin, D., Crossa, J., Lacaze, X., Du Cheyron, P., Daucourt, J., Lorgeou, J., & de los Campos, G. (2014). A reaction norm model for genomic selection using high-dimensional genomic and environmental data. *Theoretical and Applied Genetics*, 127(3), 595–607. <https://doi.org/10.1007/s00122-013-2243-1>
- Kanwar, R. S., Baker, J. L., & Mukhtar, S. (1988). Excessive soil-water effects at various stages of development on the growth and yield of corn. *Transactions of the ASAE*, 31(1), 133–141. <https://doi.org/10.13031/2013.30678>
- Kaur, G., Singh, G., Motavalli, P. P., Nelson, K. A., Orlowski, J. M., & Golden, B. R. (2019). Impacts and management strategies for

- crop production in waterlogged or flooded soils: A review. *Agronomy Journal*, 112(3), 1475–1501. <https://doi.org/10.1002/agj2.20093>
- Keating, B. A., Carberry, P. S., Hammer, G. L., Probert, M. E., Robertson, M. J., Holzworth, D., & Smith, C. J. (2003). An overview of APSIM, a model designed for farming systems simulation. *European Journal of Agronomy*, 18(3–4), 267–288. [https://doi.org/10.1016/S1161-0301\(02\)00108-9](https://doi.org/10.1016/S1161-0301(02)00108-9)
- Kettlewell, P. S., Sothorn, R. B., & Koukkari, W. L. (1999). UK wheat quality and economic value are dependent on the north Atlantic oscillation. *Journal of Cereal Science*, 29(3), 205–209. <https://doi.org/10.1006/jcrs.1999.0258>
- Kheir, A. M. S., Alkharabsheh, H. M., Seleiman, M. F., Al-Saif, A. M., Ammar, K. A., Attia, A., Zoghdan, M. G., Shabana, M. M. A., Aboelsoud, H., & Schillaci, C. (2021). Calibration and validation of AQUACROP and APSIM models to optimize wheat yield and water saving in arid regions. *Land*, 10(12), 1375. <https://doi.org/10.3390/land10121375>
- Kusmec, A., Zheng, Z. H., Archontoulis, S., Ganapathysubramanian, B., Hu, G. P., Wang, L. Z., & Schnable, P. S. (2021). Interdisciplinary strategies to enable data-driven plant breeding in a changing climate. *One Earth*, 4(3), 372–383. <https://doi.org/10.1016/j.oneear.2021.02.005>
- Lawrence-Dill, C. J., Schnable, P. S., & Springer, N. M. (2019). Idea factory: The maize genomes to fields initiative. *Crop Science*, 59(4), 1406–1410. <https://doi.org/10.2135/cropsci2019.02.0071>
- Legates, D. R., & McCabe, G. J. (1999). Evaluating the use of “goodness-of-fit” measures in hydrologic and hydroclimatic model validation. *Water Resources Research*, 35(1), 233–241. <https://doi.org/10.1029/1998wr900018>
- Lenhart, T., Eckhardt, K., Fohrer, N., & Frede, H. G. (2002). Comparison of two different approaches of sensitivity analysis. *Physics and Chemistry of the Earth*, 27(9–10), 645–654. [https://doi.org/10.1016/S1474-7065\(02\)00049-9](https://doi.org/10.1016/S1474-7065(02)00049-9)
- Loffler, C. M., Wei, J., Fast, T., Gogerty, J., Langton, S., Bergman, M., & Cooper, M. (2005). Classification of maize environments using crop simulation and geographic information systems. *Crop Science*, 45(5), 1708–1716. <https://doi.org/10.2135/cropsci2004.0370>
- Long, D. S., & McCallum, J. D. (2020). Adapting a relatively low-cost reflectance spectrometer for on-combine sensing of grain protein concentration. *Computers and Electronics in Agriculture*, 174, 105467. <https://doi.org/10.1016/j.compag.2020.105467>
- Malosetti, M., Bustos-Korts, D., Boer, M. P., & van Eeuwijk, F. A. (2016). Predicting responses in multiple environments: Issues in relation to genotype x environment interactions. *Crop Science*, 56(5), 2210–2222. <https://doi.org/10.2135/cropsci2015.05.0311>
- Martinez-Feria, R. A., Castellano, M. J., Dietzel, R. N., Helmers, M. J., Liebman, M., Huber, I., & Archontoulis, S. V. (2018). Linking crop- and soil-based approaches to evaluate system nitrogen-use efficiency and tradeoffs. *Agriculture Ecosystems & Environment*, 256, 131–143. <https://doi.org/10.1016/j.agee.2018.01.002>
- Mavromatis, T., Boote, K. J., Jones, J. W., Irmak, A., Shinde, D., & Hoogenboom, G. (2001). Developing genetic coefficients for crop simulation models with data from crop performance trials. *Crop Science*, 41(1), 40–51. <https://doi.org/10.2135/cropsci2001.41140x>
- McFarland, B. A., AlKhalifah, N., Bohn, M., Bubert, J., Buckler, E. S., Ciampitti, I., & de Leon, N. (2020). Maize genomes to fields (G2F): 2014–2017 field seasons: Genotype, phenotype, climatic, soil, and inbred ear image datasets. *BMC Research Notes*, 13(1), 71. <https://doi.org/10.1186/s13104-020-4922-8>
- Messina, C. D., Podlich, D., Dong, Z. S., Samples, M., & Cooper, M. (2011). Yield-trait performance landscapes: From theory to application in breeding maize for drought tolerance. *Journal of Experimental Botany*, 62(3), 855–868. <https://doi.org/10.1093/jxb/erq329>
- Messina, C. D., Technow, F., Tang, T., Totir, R., Ghossein, 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*, 100, 151–162. <https://doi.org/10.1016/j.eja.2018.01.007>
- Mikel, M. A., & Dudley, J. W. (2006). Evolution of North American dent corn from public to proprietary germplasm. *Crop Science*, 46(3), 1193–1205. <https://doi.org/10.2135/cropsci2005.10-0371>
- Mueller, N. D., Gerber, J. S., Johnston, M., Ray, D. K., Ramankutty, N., & Foley, J. A. (2012). Closing yield gaps through nutrient and water management. *Nature*, 490(7419), 254–257. <https://doi.org/10.1038/nature11420>
- Mukhtar, S., Baker, J. L., & Kanwar, R. S. (1990). Corn growth as affected by excess soil-water. *Transactions of the ASAE*, 33(2), 437–442. <https://doi.org/10.13031/2013.31348>
- Ordóñez, R. A., Castellano, M. J., Hatfield, J. L., Helmers, M. J., Licht, M. A., Liebman, M., & Archontoulis, S. V. (2018). Maize and soybean root front velocity and maximum depth in Iowa, USA. *Field Crops Research*, 215, 122–131. <https://doi.org/10.1016/j.fcr.2017.09.003>
- Ordóñez, R. A., Savin, R., & Slafer, G. A. (2015). Genetic variation in the critical specific leaf nitrogen maximising yield among modern maize hybrids. *Field Crops Research*, 172, 99–105. <https://doi.org/10.1016/j.fcr.2014.12.002>
- Pasley, H. R., Huber, I., Castellano, M. J., & Archontoulis, S. V. (2020). Modeling flood-induced stress in soybeans. *Frontiers in Plant Science*, 11, 62. <https://doi.org/10.3389/fpls.2020.00062>
- Perry, E. D., Yu, J., & Tack, J. (2020). Using insurance data to quantify the multidimensional impacts of warming temperatures on yield risk. *Nature Communications*, 11(1), 4542. <https://doi.org/10.1038/s41467-020-17707-2>
- Probert, M. E., Dimes, J. P., Keating, B. A., Dalal, R. C., & Strong, W. M. (1998). APSIM’s water and nitrogen modules and simulation of the dynamics of water and nitrogen in fallow systems. *Agricultural Systems*, 56(1), 1–28. [https://doi.org/10.1016/S0308-521x\(97\)00028-0](https://doi.org/10.1016/S0308-521x(97)00028-0)
- Puntel, L. A., Sawyer, J. E., Barker, D. W., Dietzel, R., Poffenbarger, H., Castellano, M. J., & Archontoulis, S. V. (2016). Modeling long-term corn yield response to nitrogen rate and crop rotation. *Frontiers in Plant Science*, 7, 1630. <https://doi.org/10.3389/fpls.2016.01630>
- R Development Core Team. (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>
- Ray, D. K., Mueller, N. D., West, P. C., & Foley, J. A. (2013). Yield trends are insufficient to double global crop production by 2050. *PLoS One*, 8(6), e66428. <https://doi.org/10.1371/journal.pone.0066428>
- Ren, B. Z., Zhang, J., Dong, S., Liu, P., & Zhao, B. (2016a). Effects of duration of waterlogging at different growth stages on grain growth of summer maize (*Zea mays* L.) under field conditions. *Journal of Agronomy and Crop Science*, 202(6), 564–575. <https://doi.org/10.1111/jac.12183>
- Ren, B. Z., Zhang, J. W., Dong, S. T., Liu, P., & Zhao, B. (2016b). Effects of waterlogging on leaf mesophyll cell ultrastructure and

- photosynthetic characteristics of summer maize. *PLoS One*, 11(9), e0161424. <https://doi.org/10.1371/journal.pone.0161424>
- Ren, B. Z., Zhang, J. W., Dong, S. T., Liu, P., & Zhao, B. (2016c). Root and shoot responses of summer maize to waterlogging at different stages. *Agronomy Journal*, 108(3), 1060–1069. <https://doi.org/10.2134/agnonj2015.0547>
- Ren, B. Z., Zhang, J. W., Li, X., Fan, X., Dong, S. T., Liu, P., & Zhao, B. (2014). Effects of waterlogging on the yield and growth of summer maize under field conditions. *Canadian Journal of Plant Science*, 94(1), 23–31. <https://doi.org/10.4141/Cjps2013-175>
- Reynolds, M., Chapman, S., Crespo-Herrera, L., Molero, G., Mondal, S., Pequeno, D. N. L., & Sukumaran, S. (2020). Breeder friendly phenotyping. *Plant Science*, 295, 110396. <https://doi.org/10.1016/j.plantsci.2019.110396>
- Romay, M. C., Millard, M. J., Glaubitz, J. C., Peiffer, J. A., Swarts, K. L., Casstevens, T. M., & Gardner, C. A. (2013). Comprehensive genotyping of the USA national maize inbred seed bank. *Genome Biology*, 14(6), R55. <https://doi.org/10.1186/gb-2013-14-6-r55>
- Schnable, P. S., Ware, D., Fulton, R. S., Stein, J. C., Wei, F. S., Pasternak, S., & Wilson, R. K. (2009). The b73 maize genome: Complexity, diversity, and dynamics. *Science*, 326(5956), 1112–1115. <https://doi.org/10.1126/science.1178534>
- Schulz-Streeck, T., Ogotu, J. O., Gordillo, A., Karaman, Z., Knaak, C., & Piepho, H. P. (2013). Genomic selection allowing for marker-by-environment interaction. *Plant Breeding*, 132(6), 532–538. <https://doi.org/10.1111/pbr.12105>
- Seidel, S. J., Palosuo, T., Thorburn, P., & Wallach, D. (2018). Towards improved calibration of crop models—Where are we now and where should we go? *European Journal of Agronomy*, 94, 25–35. <https://doi.org/10.1016/j.eja.2018.01.006>
- Sekhon, R. S., Joyner, C. N., Ackerman, A. J., McMahan, C. S., Cook, D. D., & Robertson, D. J. (2020). Stalk bending strength is strongly associated with maize stalk lodging incidence across multiple environments. *Field Crops Research*, 249, 107737. <https://doi.org/10.1016/j.fcr.2020.107737>
- Shorter, R., Lawn, R. J., & Hammer, G. L. (1991). Improving genotypic adaptation in crops—A role for breeders, physiologists and modelers. *Experimental Agriculture*, 27(2), 155–175. <https://doi.org/10.1017/S0014479700018810>
- So, Y. S., & Edwards, J. (2009). A comparison of mixed-model analyses of the Iowa crop performance test for corn. *Crop Science*, 49(5), 1593–1601. <https://doi.org/10.2135/cropsci2008.09.0574>
- Soufizadeh, S., Munaro, E., McLean, G., Massignam, A., van Oosterom, E. J., Chapman, S. C., & Hammer, G. L. (2018). Modelling the nitrogen dynamics of maize crops—Enhancing the APSIM maize model. *European Journal of Agronomy*, 100, 118–131. <https://doi.org/10.1016/j.eja.2017.12.007>
- Struik, P. C., Cassman, K. G., & Koornneef, M. (2007). A dialogue on interdisciplinary collaboration to bridge the gap between plant genomics and crop sciences. In *Scale and complexity in plant systems research: Gene-plant-crop relations* (vol. 21, pp. 319–328). Springer.
- Technow, F., Messina, C. D., Totir, L. R., & Cooper, M. (2015). Integrating crop growth models with whole genome prediction through approximate Bayesian computation. *PLoS One*, 10(6), e0130855. <https://doi.org/10.1371/journal.pone.0130855>
- Tian, L. X., Bi, W. S., Ren, X. S., Li, W. L., Sun, L., & Li, J. (2020). Flooding has more adverse effects on the stem structure and yield of spring maize (*Zea mays* L.) than waterlogging in northeast China. *European Journal of Agronomy*, 117, 126054. <https://doi.org/10.1016/j.eja.2020.126054>
- Tollenaar, M., & Aguilera, A. (1992). Radiation use efficiency of an old and a new maize hybrid. *Agronomy Journal*, 84(3), 536–541. <https://doi.org/10.2134/agnonj1992.00021962008400030033x>
- Tollenaar, M., Muldoon, J. F., & Daynard, T. B. (1984). Differences in rates of leaf appearance among maize hybrids and phases of development. *Canadian Journal of Plant Science*, 64(3), 759–763. <https://doi.org/10.4141/cjps84-104>
- van Eeuwijk, F. A., Bustos-Korts, D., Millet, E. J., Boer, M. P., Kruijer, W., Thompson, A., & Chapman, S. C. (2019). Modelling strategies for assessing and increasing the effectiveness of new phenotyping techniques in plant breeding. *Plant Science*, 282, 23–39. <https://doi.org/10.1016/j.plantsci.2018.06.018>
- Wallach, D., Palosuo, T., Thorburn, P., Hochman, Z., Gourdain, E., Andrianasolo, F., & Seidel, S. J. (2021). The chaos in calibrating crop models: Lessons learned from a multi-model calibration exercise. *Environmental Modelling and Software*. <https://doi.org/10.1016/j.envsoft.2021.105206>
- Wang, E. L., Brown, H. E., Rebetzke, G. J., Zhao, Z. G., Zheng, B. Y., & Chapman, S. C. (2019). Improving process-based crop models to better capture genotype x environment x management interactions. *Journal of Experimental Botany*, 70(9), 2389–2401. <https://doi.org/10.1093/jxb/erz092>
- Wickham, H. (2017). Tidyverse: Easily install and load the ‘tidyverse’. R package (Version 1.2.1). <https://CRAN.R-project.org/package=tidyverse>
- Wickham, H., & Bryan, J. (2019). Readxl: Read excel files. R package (Version 1.3.1). <https://CRAN.R-project.org/package=readxl>
- Wickham, H., François, R., Henry, L., & Müller, K. (2019). Dplyr: A grammar of data manipulation. R package (Version 0.8.3). <https://CRAN.R-project.org/package=dplyr>
- Willmott, C. J. (1981). On the validation of models. *Physical Geography*, 2, 184–194.
- Yan, P., Yue, S. C., Qiu, M. L., Chen, X. P., Cui, Z. L., & Chen, F. J. (2014). Using maize hybrids and in-season nitrogen management to improve grain yield and grain nitrogen concentrations. *Field Crops Research*, 166, 38–45. <https://doi.org/10.1016/j.fcr.2014.06.012>
- Yang, H. S., Dobermann, A., Lindquist, J. L., Walters, D. T., Arkebauer, T. J., & Cassman, K. G. (2004). Hybrid-maize—A maize simulation model that combines two crop modeling approaches. *Field Crops Research*, 87(2–3), 131–154. <https://doi.org/10.1016/j.fcr.2003.10.003>
- Yin, X., Struik, P. C., Gu, J., & Wang, H. (2016). Modelling qtl-trait-crop relationships: Past experiences and future prospects. In X. Yin & P. Struik (Eds.), *Crop systems biology* (pp. 193–218). Springer. https://doi.org/10.1007/978-3-319-20562-5_9
- Zaidi, P. H., Rafique, S., Rai, P. K., Singh, N. N., & Srinivasan, G. (2004). Tolerance to excess moisture in maize (*Zea mays* L.): Susceptible crop stages and identification of tolerant genotypes. *Field Crops Research*, 90(2–3), 189–202. <https://doi.org/10.1016/j.fcr.2004.03.002>
- Zambrano-Bigiarini, M. (2020). Hydrogof: Goodness-of-fit functions for comparison of simulated and observed hydrological time series. R package (Version 0.4-0). <https://github.com/hzambran/hydroGOF>
- Zhang, L., Liang, Z. Y., He, X. M., Meng, Q. F., Hu, Y. C., Schmidhalter, U., & Chen, X. P. (2020). Improving grain yield and protein concentration of maize (*Zea mays* L.) simultaneously by appropriate

hybrid selection and nitrogen management. *Field Crops Research*, 249, 107754. <https://doi.org/10.1016/j.fcr.2020.107754>

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Winn, C. A., Archontoulis, S., & Edwards, J. (2023). Calibration of a crop growth model in APSIM for 15 publicly available corn hybrids in North America. *Crop Science*, 63, 511–534. <https://doi.org/10.1002/csc2.20857>