Agricultural Systems 150 (2017) 109–119

Contents lists available at ScienceDirect

Agricultural Systems

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

A comparison of the performance of the CSM-CERES-Maize and EPIC models using maize variety trial data

Yawen Bao [a], Gerrit Hoogenboom [a,b,d,*], Ron McClendon [a], George Vellidis [c]

[a] College of Engineering, The University of Georgia, Athens, GA 30602, USA
[b] AgWeatherNet Program, Washington State University, Prosser, WA 99350-8694, USA
[c] Engineering Building, The University of Georgia, Tifton Campus, 2329 Rainwater Rd, Tifton, GA 31793, USA
[d] Institute for Sustainable Food Systems, Frazier Rogers Hall, University of Florida, Gainesville, FL 32611, USA

ARTICLE INFO

Article history:
Received 24 January 2016
Received in revised form 11 October 2016
Accepted 16 October 2016
Available online xxxx

Keywords:
Yield
Calibration
Evaluation
Climate change
Decision support system

ABSTRACT

Multiple crop models are now being used in climate change impact studies. However, calibration of these models with local data is still important, but often this information is not available. This study determined the feasibility of using maize variety trial data for the evaluation of the CSM-CERES-Maize and EPIC models. The models were calibrated using observed grain yield from variety trials conducted in Blairsville, Calhoun, Griffin, Midville, Plains, and Tifton, Georgia, USA. The software program GenCALC was used to calibrate the yield component coefficients of CSM-CERES-Maize, while the coefficients for EPIC were manually adjusted. The criteria for evaluating the performance of the two crop models included the slope of linear regression, R^2, d-stat, and RMSE. Following model calibration and evaluation, both models were used to simulate rainfed and irrigated grain yield during 1958 to 2012 for the same six locations that were used for model evaluation. The differences between the simulations of CSM-CERES-Maize and observations were no more than 3% for calibration and no more than 8% for evaluation. However, the differences between the simulations of EPIC and observations ranged from 2% to 23% for calibration and evaluation, which was larger than for the CSM-CERES-Maize model. This analysis showed that calibration of CSM-CERES-Maize was slightly superior than EPIC for some cultivars. Although this study only used observed grain yield for calibration and evaluation, the results showed that both calibrated models can provide fairly accurate simulations. Therefore, it can be concluded that limited data sets from maize variety trials can be used for model calibration when detailed data from growth analysis studies are not readily available.

1. Introduction

"Crop simulation models integrate the current state-of-the art scientific knowledge from many different disciplines, including crop physiology, plant breeding, agronomy, agrometeorology, soil physics, soil chemistry, soil fertility, plant pathology, entomology, economics and many others" (Hoogenboom, 2000). Since agricultural production is determined by weather and climate (Adams et al., 1998), these models have been used extensively to analyze the potential impact of climate change on crop production (Lobell and Asner, 2003; Semenov and Shewry, 2011; White and Hoogenboom, 2010). Coupling crop models and climate models has been widely used in both past and current climate impact analysis (Carbone et al., 2003; Curry et al., 1995; Easterling et al., 1996; Easterling et al., 1997; Parry et al., 2004; Parry et al., 2007; White et al., 2011). Alexandrov and Hoogenboom (2000) combined the CERES v.3.5 simulation model for maize (*Zea mays* L.) and winter wheat (*Triticum aestivum* L.) and the CROPGRO v.3.5 model for soybean (*Glycine* max L.) and peanut (*Arachis hypogaea* L.) with climate projections of Global Circulation Models (GCM) for more than 500 locations in the southeastern region of the USA. Their results concluded that the GCM scenarios projected a decrease in crop yield for the 2020s under the current level of CO_2 and the increased CO_2 tended to increase crop yield. Adaptation options were suggested for changing sowing date, hybrids and cultivar selection, and fertilization to mitigate the potential negative impact of potential warming.

It is well known that the calibration and evaluation of a crop model is extremely important when a crop model is applied for new locations with new varieties, cultivars or hybrids. Model evaluation is not only important for determining the accuracy of the simulations, such as for flowering, maturity and yield, but also to show the possible uncertainties that a crop model could introduce in impact studies. Many studies have developed procedures for the calibration of crop models based on limited observations for numerous applications for a range of crops such as maize, soybean, alfalfa (*Medicago sativa*), grain sorghum (*Sorghum bicolor* (L.) *Moench*), wheat, barley (*Hordeum vulgare* L.), peanut, rice (*Oryza sativa*), cotton (*Gossypium hirsutum L.*), etc. (Balkovič et al.,

* Corresponding author at: Institute for Sustainable Food Systems, Frazier Rogers Hall, University of Florida, Gainesville, FL 32611, USA.
E-mail address: gerrit@ufl.edu (G. Hoogenboom).

http://dx.doi.org/10.1016/j.agsy.2016.10.006
0308-521X/

2013; Cabelguenne et al., 1990; Gaiser et al., 2010; Ko et al., 2009; Perez-Quezada et al., 2003; Soler et al., 2007).

In addition to the calibration and evaluation of single model, studies also have shown that different modeling approaches may lead to significant differences in results due to the differences between crop simulation models (Wolf, 2002). The comparison of the performance of different crop models in predicting crop phenology has been studied (Porter et al., 1993, and French and Hodges, 1985) and for grain yield (e.g., Cerrato and Blackmer, 1990), showing that some models performed better than others, which means less uncertainties will be introduced when the models are applied. Recent discussion of uncertainties that crop models could introduce in climate change impact studies emphasizes a comparison of the performance of different crop models (Ceglar et al., 2011; Rötter et al., 2012; Semenov and Stratonovitch, 2010). Newly released cultivars, varieties, and hybrids have not been parameterized for most models and, therefore, need to be calibrated, while the crop models also have improved over time (Holzworth et al., 2015). Therefore, the comparison of the performance among different crop models and the use of multiple crop models to minimize uncertainties has been acted on internationally, such as in The Agricultural Model Intercomparison and Improvement Project (Rosenzweig et al., 2013). In addition to calibration and evaluation of each model, a proper sensitivity test is also important in order to better understand the potential impact of climate change effect on crop growth, development and ultimately yield.

Comprehensive data sets and associated data standards are needed for the comparison of crop models' performance, especially for the more complex dynamic crop growth simulation models (Hunt et al., 2001; Hoogenboom et al., 2012a; White et al., 2013). For instance, Anothai et al. (2008) collected detailed phenological and growth analysis data for the calibration of CSM-CROPGRO-Peanut. However, detailed growth analysis data are normally not available and are also very expensive to obtain with respect to financial resources required for field experimentation and personnel resources for detailed data collection (Kersebaum et al., 2015). Unfortunately for most impact studies, the calibration and evaluation procedures of the crop simulation models have been ignored, and the recommended cultivar coefficients from model designers or previous studies were used, introduction additional uncertainties.

Only a few studies so far have concentrated on multiple model comparisons, such as for barley (Rötter et al., 2012), wheat (Asseng et al., 2013; Liu et al., 2016), maize (Bassu et al., 2014) and potato (Fleisher et al., 2016). There is, therefore, also a need to analyze the uncertainties of maize crop models with recently released maize hybrids. In this study two commonly used maize crop simulation models in both the USA and across the globe were selected. One is CSM-CERES-Maize, which is one module of the Decision Support System for Agrotechnology Transfer (DSSAT), the other one is Environmental Policy Integrated Climate (EPIC) cropping systems model. As defined by White and Hoogenboom (2003), EPIC can be considered a type 2 model with species-specific genetic coefficients but no reference to genotypes, while CSM-CERES-Maize is a type 3 model with genotypic differences represented by cultivar-specific genetic coefficients. The main interest in this study was to compare two models with different sets of genetic coefficients rather than the performance of an ensemble requiring more than two models.

DSSAT is a software package that incorporates independent models for more than 25 different crops with programs that facilitate the evaluation and application of the crop models for different purposes (Hoogenboom et al., 2012b; Jones et al., 2003). The DSSAT crop models simulate growth, development, and yield by considering weather, genetics, soil water, soil carbon and nitrogen, and management for single or multiple seasons and in crop rotations at any location where minimum inputs are provided (Hunt and Boote, 1998; Jones et al., 2003). The minimum inputs contain soil profile, daily weather data (minimum and maximum temperature, precipitation, and solar radiation), crop management (plant population, row spacing, application of irrigation and fertilizer etc.), and a set of cultivar coefficients. The individual crop growth modules of CSM such as CERES and CROPGRO were designed for simulating different crops to provide an accurate description for the development stages of a specific cultivar. The CSM-CERES-Maize is the module that simulates growth, development and yield for maize using a daily time step. Growth stages that are simulated by CSM-CERES-Maize include germination, emergence, end of juvenile, floral induction, 75% silking, beginning grain fill, maturity, and harvest (Jones and Kiniry, 1986; Jones et al., 2003; Ritchie et al., 1998). The physiological day accumulator is a function of temperature and day length; when it reaches the threshold given in the cultivar file, the new growth stages is triggered. The potential growth depends on photosynthetically active radiation and its interception, where the actual biomass production is constrained by stresses such as temperature, nitrogen, and water. It also considers the sensitivity of a crop to the ambient CO_2 concentration.

EPIC was designed to estimate soil productivity as affected by erosion throughout the U.S. (Williams et al., 1989). The components of the EPIC model include weather, hydrology, erosion-sedimentation, nutrient cycling, crop growth, tillage, soil temperature, economics, and plant environment control (Jones et al., 1984a, 1984b; Sharpley et al., 1984; Williams et al., 1984, 1989). Similar to CSM-CERES-Maize, soil profile information, daily weather data, crop management, and a set of cultivar coefficients are the minimum data inputs for EPIC. However, multiple crops are simulated by a single module. The yield is estimated using the harvest index and above-ground biomass. The above-ground biomass in turn is a function of photosynthetically active radiation and leaf area. Leaf area is calculated as a function of heat unit accumulation, crop development states and crop stresses. Unfortunately, this model does not provide the individual predictions and thus outputs for crop development stages.

The goal of this study was to determine the feasibility of using limited maize variety trial data for the evaluation of different crop simulation models using different complexities with respect to genetic coefficients. The first objective was to determine the cultivar coefficients for the two crop models using observed grain yield; the second objective was to determine whether the performance of the two evaluated crop models is comparable in predicting maize grain yield.

2. Materials and methods

2.1. Experimental data collection

In Georgia, variety trials for both rainfed and irrigated maize are conducted at the regional agricultural experimental stations located in Blairsville (34.84°N, 83.93°W), Calhoun (34.34°N, 85.12°W), Griffin (33.26°N, 84.28°W), Midville (32.88°N, 82.22°W), Plains (32.05°N, 84.37°W), and Tifton (31.49°N, 83.53°W) (Table 1). These variety trials are conducted by the University of Georgia (UGA) College of Agricultural & Environmental Science (CAES) Statewide Variety Testing (SWVT) program. In this study data collected from 2003 until 2010 were used (Coy et al., 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010). Soil profile and soil surface data and generic soil information for these seven locations were obtained from the soil analyses conducted by Perkins et al. (1986, 1979, 1978, 1982, 1983, 1985) and Natural Resources Conservation Service (NRCS) of United States Department of Agriculture (USDA). The soil types were a Bradson clay loam for Blairsville; a Waynesboro loam, an Ethowah loam, a Rome gravelly clay loam, and a Savannah loam for Calhoun; a Pacolet sandy loam and a Cecil sandy loam for Griffin; a Tifton loamy sand and a Dothan loamy sand for Midville; a Faceville sandy loam and a Greensville sandy loam for Plains; and a Tifton loamy sand, a Fuquay loamy sand, and a Dothan loamy sand for Tifton. A soil utility program of DSSAT, SBuild, was used to create the soil inputs based on these local soil profile data.

The daily solar radiation, maximum and minimum air temperature, and precipitation for each location were obtained from the Georgia

Table 1
Maximum and minimum temperature and precipitation during the crop growing season from 2003 to 2010 for the six locations of this study. The crop growing season ranged from April to October for Blairsville, April to Sep. Calhoun, Griffin, Midville, and March to Sep. for Plains and Tifton.

		Maximum temperature (°C)			Minimum temperature (°C)			
Location	Year	Max	Min	Average	Max	Min	Average	Precipitation (mm)
Blairsville	2003	31.8	9.6	24.8	19.9	−0.9	12.2	1037
	2005	34.7	8.4	25.5	21.1	−3.8	12.5	837
	2006	34.2	7.7	25.7	21.4	−4.3	11.9	736
	2007	35.9	3.7	26.5	19.8	−5.6	12	576
	2008	28.6	7.8	24.6	21.8	−3.9	11.9	438
	2009	28.6	4.1	24.1	20	−4.9	12.6	1036
	2010	33.5	14.3	26.4	21.9	−1.3	13	812
Calhoun	2003	34.1	8.5	27.9	21.5	−0.9	15.3	964
	2004	35.3	14.7	28.2	22.5	−1.1	15.4	823
	2005	36.1	10.7	28.6	22.6	−1.7	15.2	723
	2006	38.6	18.1	29.8	22.8	−0.4	15.3	469
	2007	39.9	6.7	30.1	22.4	−6	14.6	293
	2008	37.1	10	28.7	22.7	−2.1	14.7	503
	2009	36.1	8	27.8	21.6	−4.3	15.1	675
	2010	37.4	17	30.2	23	0.5	15.7	523
Griffin	2003	32.8	7.3	27.5	22.5	4.1	16.9	954
	2004	34.8	14.4	28.2	22.4	1.3	17.2	877
	2005	35.5	13.8	27.9	24.3	1.5	17.1	867
	2006	36.7	17.7	29.4	24.1	4.3	17.4	383
	2007	38.6	7.7	29.1	25.8	−2.8	17.2	379
	2008	35.9	10.2	28.5	22.9	1.4	17.1	470
	2009	35.5	7.9	27.9	24.4	−0.4	17.6	516
	2010	37.2	17.1	30.3	25.2	4.8	18.8	546
Midville	2003	34.5	9	28.9	23.8	2.1	18.5	941
	2004	37.1	17	30.1	23.9	2.2	18.5	806
	2005	36.9	15.5	29.9	25.3	4.3	18.3	614
	2006	38.4	17.8	30.8	24.4	3.6	18.3	359
	2007	39.5	11	30.7	25.4	−1.5	17.8	475
	2008	38.1	14	30.4	24.2	1.9	18.3	494
	2009	37	9.9	30	26.2	1.9	18.7	824
	2010	38.5	20.3	31.9	25.8	6	19.3	539
Plains	2003	34.6	8.9	28.1	23.1	−0.7	16.9	846
	2004	36.2	14.9	28.8	23.6	0	16.6	866
	2005	36.2	6.4	27.9	24.9	−2.8	16.4	1084
	2006	38.8	14.2	29.7	24	−0.1	16.7	687
	2007	39.2	11	29.7	24.6	−1.1	16.3	535
	2008	37.4	10.5	28.4	23	−2	16	704
	2009	36	8.9	27.7	24.6	−3.7	16.5	858
	2010	38.8	10.5	29.8	25.5	−1.4	17.4	568
Tifton	2003	34.4	10.9	28.2	23.6	0.5	18.2	987
	2004	35.1	14.6	28.8	25.5	2	18.1	939
	2005	35	7.5	27.8	25.2	−2.3	17.6	781
	2006	36.5	13.3	29.4	25	1.1	17.7	421
	2007	37.3	11.8	29.3	25.4	0.1	17.5	537
	2008	35.4	11.3	28.4	24.2	−0.1	17.5	663
	2009	35.8	9	28.3	25	−1.9	18.1	1054
	2010	37.5	11.3	29.4	25.4	−0.8	18.3	648

Automated Environmental Monitoring Network (GAEMN, www.georgiaweather.net), which was first deployed in 1991 (Hoogenboom, 1996), with 60 operational stations in 2004 (Garcia y Garcia and Hoogenboom, 2005) and over 80 in 2013. The typical maize growing season ranges from April until October for Blairsville, from April until September for Calhoun, Griffin, and Midville, and from March until September for Plains and Tifton. Blairsville has the highest latitude and elevation and, therefore, has a relatively longer growing season than the other locations, while Tifton, located in the Coastal Plains, has the lowest latitude and elevation. Precipitation varied among locations and among years due to the variable summer thunderstorms that normally occur in Georgia. Some of the locations had a dry season, defined as less than 400 mm, including Calhoun in 2007, Griffin in 2006 and 2007, and Midville in 2006 (Table 1).

Crop management, planting dates, irrigation amount, fertilizer amount, and planting population corresponded to the local management of the variety trials. Plant population at seeding was around 6 to 8 plants/m^2, row spacing was 76 cm, and the planting depth was 5 cm. The reported dates and amount of irrigation for each individual trial were also obtained and the irrigation method was sprinkler irrigation. Previous crops grown in these fields included maize, cotton, soybean, and peanut, while in some instances there was a fallow season.

The hybrids, Dyna-Gro V5373VT3, Pioneer 33M57(Hx1/LL/RR2), SS 731CL, Croplan Genetics 851 VT3 PRO, Croplan Genetics 8756 VT3, DeKalb DKC69-71(RR2/YGCB), and Pioneer 31D58 were selected because these were grown in all locations from 2003 until 2010 (Table 2). The observations included grain yield at 15.5% moisture and final harvest dates, which were used for model calibration and evaluation. Observed grain yield was corrected to 0% water content because the crop models only predict dry grain yield.

2.2. Calibration and evaluation

2.2.1. CSM-CERES-Maize

Model calibration and evaluation were based on comparing the model simulations with observations. Multiple years (2003 to 2010) were considered with some used for calibration and the rest for evaluation (Table 2). The cultivar coefficients were adjusted in order for the simulated variables to fit the observations. The cultivar coefficients of the CSM-CERES-Maize model include thermal time from seedling emergence to the end of the juvenile phase (P1), extent to which development is delayed for each hour increase in photoperiod above the longest photoperiod at which development proceeds at a maximum rate (P2), thermal time from silking to physiological maturity (P5), maximum possible number of kernels per plant (G2), kernel filling rate during the linear grain filling state and under optimum conditions (G3), and the interval in thermal time (degree days) between successive leaf tip appearances (PHINT) (Table 3). The soil fertility factor (SLPF) was also adjusted as it is an input parameter that affects the overall growth rate of simulated total biomass by modifying daily canopy photosynthesis and is attributed to soil fertility differences and soil-based pests, such as nematodes (Guerra et al., 2008; Mavromatis et al., 2001).

The calibration procedure was similar to the one developed for the CSM-CROPGRO-Soybean models (Bao et al., 2015). This included the Genotype Coefficient Calculator (GENCALC) to calibrate the parameters with corresponding observations and to manually adjust the remainder of the coefficients. GENCALC was designed for the calibration of the cultivar coefficients of DSSAT. It starts with the initial coefficients that are extracted from the genotype file of DSSAT and it selects the best value for each coefficient by evaluating the root mean square error (RMSE) between the simulated and observed variables (Hunt et al., 1993). The search for the appropriate value for each of the genetic coefficients is limited in range by setting the change for each step, i.e., STEP, and the number of times GENCALC should change the values of a particular coefficient, i.e., LOOP.

First of all, SLPF was manually adjusted for each location based on the initial set of cultivar coefficients. The values of SLPF range from 0.7 to 0.94 (Jones et al., 1989; Mavromatis et al., 2001). The adjustment started with an initial value, 0.8, until the simulated grain yield was similar to the observed grain yield. All seven hybrids for all years (2003 to 2010) were used for each of the six locations. The next step was to calibrate the cultivar coefficients. Because grain yield was only available for the variety trial data, the cultivar coefficients G2 and G3 could be automatically calibrated by using GENCALC. At the same time the cultivar coefficients P1, P2, P5, and PHINT were manually changed with a certain percentage while GENCALC optimized for G2 and G3. A sensitivity test showed that the loop for manually modifying the parameters was 10 for P1, 0.3 for P2, 10 for P5, and 1 for PHINT. The search for P1 ranged from 110 to 458, for P2 ranged from 0 to 3, for P5 ranged from 390 to 1000, and for PHINT ranged from 30 to 75. The initial values were 200, 0.3, 800, and 38.9 for P1, P2, P5, and PHINT respectively. Ideally, the simulated days from planting to maturity (maturity days) should have a

Table 2
Average grain yield for seven selected maize hybrids for six locations in Georgia.

Variety	Average grain yield (kg/ha)		Calibration years	Evaluation years
	Irrigated	Rainfed		
Dyna-Gro V5373VT3	10,400	8669	2008, 2010	2009
Pioneer 33M57 (Hx1/LL/RR2)	10,258	9183	2007, 2009	2008
SS 731CL	9582	8268	2007, 2009	2008
Croplan Genetics 851 VT3 PRO	10,470	8351	2008, 2010	2009
Croplan Genetics 8756 VT3	10,877	7908	2009, 2010	2008
DeKalb DKC69-71 (RR2/YGCB)	10,538	8807	2004, 2006, 2007, 2008, 2010	2003, 2005, 2009
Pioneer 31D58	11,619	7966	2006, 2008, 2010	2007, 2009

good fit with the observed maturity days when adjusting P1, P2, P5, and PHINT. However, because no observed maturity days were obtained, the observed days from planting to harvest (harvest days) were used, which is usually longer than the number of days to maturity. GENCALC searches G2 and G3 by comparing simulated grain yield with observations. For G2 the range was 248 to 990 and for G3 the range was 4.4 to 16.5. The initial value for G2 was 770 and 8.5 for G3. The final step was to use the calibrated cultivar coefficients for evaluation using an independent data set from the variety trial data (Table 2).

2.2.2. EPIC

EPIC also requires a number of crop-specific coefficients (Table 3), which are similar to the CSM-CERES-Maize model. The parameters that were calibrated in this study also were selected for calibration in previous studies, such as Williams et al. (1989), Cabelguenne et al. (1990) and Guerra et al. (2004), and Ko et al. (2009). The potential heat units (PHU) for maize is defined as the total number of heat units from planting to physiological maturity. Biomass-energy ratio (WA), maximum harvest index (HI), fraction of growing season when leaf area declines (DLAI), maximum potential leaf area index (DMLA) and drought sensitivity parameter (WSYF) were also adjusted. Batch processing was applied to search for each parameter within a certain range. A sensitivity test was first conducted to determine the optimum range for the optimization. The values for PHU ranged from 1600 to 2000 with a step of 10; the values for WA ranged from 40 to 55 with a step of 1; the values for HI ranged from 0.1 to 0.6 with a step of 0.05; the values for DMLA ranged from 2 to 6 with a step of 1; the values for DLAI ranged from 0.5 to 0.95 with a step of 0.05; and the values for WSYF ranged from 0.01 to 0.4 with a step of 0.01. Following calibration, an independent set of the variety trial data was used for model evaluation similar to the approach used for CSM-CERES-Maize (Table 2).

2.3. Statistical criteria

There has been an extensive discussion in the selection of the appropriate statistical criteria for evaluating simulation models, especially the use of root mean square error (RMSE) and mean absolute error (MAE) (Chai and Draxler, 2014). However, there is no best one among these statistical criteria and normally multiple criteria are selected based on the need of an individual study. We selected commonly used statistical criteria for evaluating crop models similar to Anothai et al. (2008), Mavromatis et al. (2001), Yang et al. (2014b), Soler et al. (2007) and others. The comparison between simulated and observed data for both calibration and evaluation was based on the following criteria: slope of the regression of simulated against observed, the coefficient of determination (R^2), index of agreement (d), and root mean square

Table 3
Cultivar coefficients for CSM-CERES-Maize model.

CSM-CERES-Maize Cultivar Coefficients		Min	Max	Initial value	Unit
P1	Thermal time from seedling emergence to the end of the juvenile phase	110	458	200	Degree days
P2	Extent to which development is delayed for each hour increase in photoperiod above the longest photoperiod at which development proceeds at a maximum rate	0	3	0.3	Day h^{-1}
P5	Thermal time from silking to physiological maturity	390	1000	800	Degree days
G2	Maximum possible number of kernels per plant	248	990	770	Kernel/plant
G3	Kernel filling rate during the linear grain filling state and under optimum conditions	4.4	16.5	8.5	Mg day^{-1}
PHINT	The interval in thermal time (degree days) between successive leaf tip appearances	30	75	38.9	Degree days
EPIC cultivar coefficients					
WA	Biomass-Energy ratio	40	55	40	
BE	Crop parameter - converts energy to biomass				$kg \cdot ha \cdot MJ^{-1} \cdot m^{-2}$
HI	Potential harvest index - ratio of crop yield to above ground biomass	0.1	0.6	0.5	
To	Optimal temperature for a crop				°C
Tb	Base temperature for a crop (plant start growing)				°C
DMLA	Maximum LAI potential for a crop	2	6	6	
DLAI	Fraction of growing season when leaf area starts declining	0.5	0.95	0.8	
HUIo	Heat unit index value when leaf area index starts declining				
ah1, ah2	Crop parameters that determine the shape of the leaf-area-index development curve				
af1, af2	Crop parameters for frost sensitivity				
Ad	Crop parameters that governs leaf area index decline rate				
ALT	Aluminum tolerance index number				
CAF	Critical aeration factor for a crop				
HMX	Maximum crop height				m
RDMX	Maximum root depth for a crop				m
WSFY	Water stress factor for adjusting harvest index				
bn1, bn2, bn3	Crop parameters for plant N concentration equation				
bp1, bp2, bp3	Crop parameters for plant P concentration equation				
PHU	Potential Heat Units	1600	2000	1800	°C

error (RMSE) (Casella and Berger, 2002; Yang et al., 2014a), which were defined as follows:

$$R^2 = 1 - \frac{\sum_i (O_i - P_i)^2}{\sum_i (O_i - \overline{O})^2}$$

$$d = 1 - \left[\frac{\sum_{i=1}^{n} (P_i - O_i)^2}{\sum_{i=1}^{n} (|P_i'| - |O_i'|)^2}\right]$$

$$d = 1 - \left[\frac{\sum_{i=1}^{n} (P_i - O_i)^2}{\sum_{i=1}^{n} (|P_i'| - |O_i'|)^2}\right]$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (P_i - O_i)^2}{n.}}$$

where n is the number of observations, P_i is the predicted value for the *ith* measurement, O_i is the observed value for the *ith* measurement, $\overline{O}$ is the mean of all observations, $P_i' = P_i - \overline{O}$, and $O_i' = O_i - \overline{O}$. For the linear regression of simulated against observed yield, slope, R^2, and d ranged from 0 to 1 and a best fit requires that they are 1 or close to 1. For RMSE, a smaller value means a better fit.

2.4. Comparison of CSM-CERES-Maize and EPIC

Following calibration and evaluation, both models were used to predict yield under both irrigated and rainfed conditions for Blairsville, Calhoun, Griffin, Midville, Plains, and Tifton using long-term historical weather data from 1958 to 2012. One of the objectives of this analysis was to determine the differences in yield prediction between the two models for different environments, but using the same crop management as was used in the variety trial data. The soil types varied with year for the variety trials, but for this analysis the most common soil type was used for each location. This included a Bradson clay loam for Blairsville, an Etowah loam for Calhoun, a Cecil sandy loam for Griffin, a Tifton loamy sand for Midville, a Greensville sandy loam for Plains, and a Tifton loamy sand for Tifton. An analysis of variance (one way ANOVA) along with a graphical analysis using box-plots was then conducted to determine whether the simulations of CSM-CERES-Maize and EPIC were significantly different. The null hypothesis here was that the simulations of two crop models do not have a significant difference. The level of $\alpha = 0.05$ (95% confidence level) was used; if value for p is smaller than a it means that there is a significant difference between the simulations of the two crop models.

3. Results

3.1. Evaluation of CSM-CERES-Maize

The calibrated value for the soil fertility factor (SLPF) was 0.8 for Blairsville, 0.76 and 0.70, 0.87, and 0.9 for Calhoun, 0.78 and 0.7 for Griffin, 0.82 and 0.85 for Midville, 0.84 and 0.73 for Plains, and 0.89, 0.9, and 0.89 for Tifton (Table 4). Some locations had multiple values for SLPF because the soil types varied by year. Since SLPF was estimated for each of the six locations and all hybrids for all years were used for calibration, the linear regression of each location was based on all hybrids. The statistical criteria that were used to determine the best value for SLPF were slope, R^2, and RMSE. The difference between simulated observed yield was 14% for Tifton, 11% for Plains, and less than 3% for the other four locations. The slope of the linear regression was low for Blairsville (0.391) and it ranged from 0.582 for Midville to 0.997 for Tifton. Blairsville also had a low value for R^2, 0.056, and the value for R^2 for the other locations ranged from 0.432 for Midville to 0.803 for Tifton. The d-value for Blairsville was 0.475, while for the other locations the d-value ranged from 0.811 to 0.932. Midville had the smallest RMSE, 920 kg/ha, while for Blairsville, Calhoun, Griffin, and Plains RMSE ranged from 1201 kg/ha to 1867 kg/ha, and Tifton had the largest RMSE at 2029 kg/ha.

Table 4
Estimation of the soil fertility factor (SLPF) for six locations and observed (Obs.) and simulated (Sim.) grain yield for CSM-CERES-Maize. Statistics include slope of regression; coefficient of determination (R^2); index of agreement (d-stat); and root mean square error (RMSE) between simulated and observed yield.

Location	SLPF	Obs. (kg/ha)	Sim. (kg/ha)	Slope	R^2	d-stat	RMSE (kg/ha)
Blairsville	0.8	13,276	12,870	0.391	0.056	0.475	1867
Calhoun	0.76, 0.7, 0.87, 0.9	8020	8260	0.713	0.732	0.914	1632
Griffin	0.78, 0.70	9014	9023	0.741	0.784	0.932	1201
Midville	0.82, 0.85	11,868	11,898	0.582	0.432	0.811	920
Plains	0.84, 0.73	9639	10,697	0.618	0.65	0.816	1718
Tifton	0.89, 0.9, 0.89	10,178	8801	0.997	0.803	0.898	2029

The phenology and growth coefficients of CSM-CERES-Maize model were calibrated for seven hybrids (Table 5). The value for the cultivar coefficient P1 ranged from 220 to 330; the value for P2 ranged from 0.9 to 1.8; the value for P5 ranged from 820 to 940; the value for PHINT ranged from 48.9 to 63.9; the value for G2 ranged from 646.8 to 954.8; and the value for G3 ranged from 10.94 to 12.64. In some cases the hybrid coefficients had the same value for different hybrids. For example, the value for P1 was the same for the hybrids Dyna-Gro V5373VT3 and Croplan Genetics 851 VT3 PRO, while for G2 the hybrids Dyna-Gro V5373VT3, Pioneer 33M57 (Hx1/LL/RR2), Croplan Genetics 851 VT3 PRO, and DeKalb DKC69-71(RR2/YGCB) had the same value.

Following calibration simulated grain yield was compared with the observed yield (Table 6). In general, the performance of the model varied among the hybrids. For the hybrids Dyna-Gro V5373VT3, Pioneer 33M57 (Hx1/LL/RR2), and SS 731CL grain yield was over-estimated, which is the expected result since the limitations for simulations are less than reality. However, for some of the hybrids grain yield was under-estimated. Fortunately, the differences between simulated and observed grain yield were no more than 3% of the observations, which means a good fit. The slopes of linear regression for the seven hybrids ranged from 0.71 (SS731CL) to 1.222 (Croplan Genetics 851 VT3 PRO). Hybrid Dyna-Gro V5373VT3 had the best value, 0.997, which is close to 1. The values for R^2 of the seven cultivars ranged from 0.67 (DeKalb DKC69-71(RR2/YGCB)) to 0.885 (Dyna-Gro V5373VT3). The values of d-stat are from 0.9 (DeKalb DKC69-71(RR2/YGCB)) to 0.969 (Dyna-Gro V5373VT3) for seven hybrids. The RMSE ranged from 1033 kg/ha (Dyna-Gro V5373VT3) to 2051 kg/ha (SS 731CL).

The evaluation of CSM-CERES-Maize was conducted by comparing simulated and observed grain yield for a different set of trial data (Table 6). Yield for the hybrids Pioneer 33M57(Hx1/LL/RR2), SS 731CL, and Croplan Genetics 8756 VT3 was over-estimated and the others were under-estimated. The difference between simulated and observed yield were less than 8% of the observed yield. The values for slope of the linear regression ranged from 0.64 (Dyna-Gro V5373VT3) to 1.18 (Pioneer 33M57(Hx1/LL/RR2)). The lowest value was 0.64 for the hybrid Dyna-Gro V5373VT3, which had the highest value for the slope for the calibration. The highest value for the slope for evaluation was 0.911 for Pioneer 31D58, which is close to 1. The value for R^2 was 0.48 for DeKalb DKC69-71(RR2/YGCB), which is low, but the value for R^2 for the other hybrids ranged from 0.703 (SS 731CL) to 0.946 (Dyna-Gro V5373VT3). The values for d-stat ranged from 0.782 (Dekalb DKC69-71 (RR2/YGCB)) to 0.966 (Pioneer 33M57 (Hx1/LL/RR2)), which were similar to the values found for calibration. The RMSE ranged from 973 kg/ha to 1,980 kg/ha. The values for RMSE for model evaluation for Pioneer 33M57 (Hx1/LL/RR2) (973 kg/ha), SS 731CL (1895 kg/ha), and Croplan Genetics 8756 VT3 (1642 kg/ha) were less that the value for RMSE found during calibration. However, the other hybrids had a larger RMSE than for calibration. In summary, the simulated grain yield of the evaluation data set showed a good agreement with observed yield and was comparable to the calibration data set, with two hybrids actually performing better for model evaluation compared to model calibration.

3.2. Evaluation of EPIC

The crop simulation model EPIC was calibrated for grain yield and yield components for the same seven hybrids (Table 5) as described

Table 5
Optimized cultivar coefficients for CSM-CERES-Maize and EPIC for the seven maize hybrids.

Parameter	Dyna-Gro V5373VT3	Pioneer 33M57 (Hx1/LL/RR2)	SS 731CL	Croplan Genetics 851 VT3 PRO	Croplan Genetics 8756 VT3	DeKalb DKC69-71 (RR2/YGCB)	Pioneer 31D58
CSM-CERES-Maize							
P1	310	260	220	310	290	330	270
P2	1.8	1.5	1.2	0.9	1.8	0.9	0.9
P5	900	940	820	820	940	840	900
G2	646.8	646.8	954.8	646.8	677.6	646.8	708.4
G3	12.43	10.94	12.64	12.64	12	12.64	11.79
PHINT	63.9	58.9	53.90	48.9	63.9	48.9	58.9
EPIC							
WA	50	50	50	50	50	50	50
HI	0.45	0.50	0.55	0.45	0.5	0.45	0.5
DLAI	0.95	0.95	0.95	0.95	0.95	0.95	0.95
WSYF	0.01	0.01	0.01	0.01	0.01	0.01	0.01
DMLA	6.0	6.0	6.0	5.0	6.0	6.0	6.0
PHU	1800	1650	1800	1800	1800	1730	1770

for CSM-CERES-Maize previously. The values for the coefficient WA was the same, i.e., 50, for all hybrids; the value for HI was 0.5, except for Croplan Genetics 851 VT3 PRO, which had a value of 0.45 for HI; the value for DLAI was 0.95 for all hybrids; the value for DMLA was 6 except for Croplan Genetics 851 VT3 PRO, which had a value of 5 for DMLA. The value for WSYF was 0.01 for all hybrids, which means that they are all very sensitive to water stress. The value for PHU was 1800 for Dyna-Gro V5373VT3, SS 731CL, and Croplan Genetics 851 VT3 PRO, 1650 for Pioneer 33M57 (Hx1/LL/RR2), 1730 for DeKalb DKC69-71 (RR2/YGCB), and 1770 for Pioneer 31D58.

The accuracy of EPIC model in predicting grain yield varied with hybrids (Table 6) and was similar to the performance of the CSM-CERES-Maize model. Average simulated grain yield was over-estimated by EPIC for all hybrids. SS 731CL overestimated yield by 23%, while for the other hybrids the yield was overestimated by 2% to 15%. The slopes of linear regression ranged from 0.514 (Pioneer 33M57 (Hx1/LL/RR2)) to 0.88 (Croplan Genetics 851 VT3 PRO), while the values for R^2 ranged from 0.54 (DeKalb DKC69-71 (RR2/YGCB)) to 0.814 (Dyn-Gro V5373VT3). The values for the d-stat ranged from 0.754 for SS 731CL to 0.947 for Dyn-Gro V5373VT3, which is close to 1. RMSE ranged from 1268 kg/ha (Croplan Genetics 851 VT3 PRO) to 2308 kg/ha (Pioneer 31D58), except for the hybrid SS 731CL with a RMSE, 3,772 kg/ha.

The evaluation of hybrids coefficients showed that EPIC over-estimated the average grain yield for all hybrids by about 10 to 23% when compared with the observations. The slopes of the linear regression were as low as 0.222 and 0.266 for the hybrids DeKalb DKC69-71 (RR2/YGCB) and Pioneer 31D58, respectively. The slopes for the other hybrids ranged from 0.555 for Dyn-Gro V5373VT3 to 1.26 for SS 731CL. The slope for Pioneer 33M57 (Hx1/LL/RR2) was 0.98, which was the best one as it was close to a perfect slope of 1. The hybrid DeKalb DKC69-71 (RR2/YGCB), not only had a low value for the slope, but also had lower values for both R^2 and d-stat, which were 0.19 and 0.575, respectively. The values for R^2 ranged from 0.49 for Pioneer 31D58 to 0.86 for Croplan Genetics 8756 VT3, while the values for d-stat ranged from 0.633 for Pioneer 31D58 to 0.875 for Dyn-Gro V5373VT3. The values for RMSE ranged from 1875 kg/ha for Pioneer 33M57 (Hx1/LL/RR2) to 4228 kg/ha for SS 731CL.

3.3. Evaluation of the ensemble simulations

In order to determine if an ensemble of two models would perform better than a single model, the simulations of CSM-CERES-Maize and EPIC were combined for both model calibration and model evaluation (Table 6). For the calibration, the simulated yield for all hybrids was overestimated. The yield for hybrid SS 731CL was overestimated by about 12%, the simulated yield for Pioneer 33M57(Hx1/LL/RR2) was overestimated by 8%, while for the other hybrids the overestimation ranged from about 1% to 4% when compared to the observed yield. In general, the combined simulations of the two crop models showed a good fit when compared with the observations. In addition, the evaluation of those hybrids also showed a relative small difference compared to the observations. The hybrids SS 731CL and Croplan Genetics 8756

Table 6
The average observed (Obs.) and simulated (Sim.) grain yield for the CSM-CERES-Maize and EPIC calibration and evaluation of the seven hybrids. Statistics include slope of regression; coefficient of determination (R^2); index of agreement (d-stat); and root mean square error (RMSE) of simulated and observed yield.

Calibration		Sim. (kg/ha)			Slope		R^2		d-stat		RMSE (kg/ha)	
Variety	Obs. (kg/ha)	CERES	EPIC	Combined Sim. (kg/ha)	CERES	EPIC	CERES	EPIC	CERES	EPIC	CERES	EPIC
Dyna-Gro V5373VT3	9891	9912	10,102	10,007	0.997	0.866	0.885	0.814	0.969	0.947	1033	1268
Pioneer 33M57 (Hx1/LL/RR2)	10,263	10,310	11,815	11,063	0.747	0.514	0.812	0.755	0.94	0.83	1512	2279
SS 731CL	9630	9725	11,937	10,831	0.710	0.600	0.715	0.587	0.909	0.754	2051	3772
Croplan Genetics 851 VT3 PRO	10,068	9846	10,459	10,153	1.222	0.880	0.803	0.713	0.921	0.909	1378	1268
Croplan Genetics 8756 VT3	10,083	10,022	10,907	10,465	0.822	0.684	0.734	0.785	0.922	0.898	1515	1602
DeKalb DKC69-71 (RR2/YGCB)	9897	9643	10,454	10,049	0.832	0.700	0.67	0.54	0.9	0.85	1683	1713
Pioneer 31D58	10,311	10,014	11,467	10,741	0.863	0.710	0.744	0.603	0.925	0.84	1644	2308
Evaluation												
Dyna-Gro V5373VT3	9649	9326	9530	9428	0.64	0.555	0.946	0.681	0.941	0.875	1436	2094
Pioneer 33M57 (Hx1/LL/RR2)	9678	9725	11,223	10,474	1.18	0.980	0.897	0.838	0.966	0.872	973	1875
SS 731CL	9128	9559	10,961	10,260	1.083	1.260	0.703	0.854	0.892	0.66	1895	4228
Croplan Genetics 851 VT3 PRO	9219	8498	10,108	9303	0.884	0.557	0.711	0.630	0.902	0.84	1980	2161
Croplan Genetics 8756 VT3	9745	10,434	11,995	11,215	0.902	1.100	0.732	0.860	0.91	0.84	1642	2569
DeKalb DKC69-71 (RR2/YGCB)	10,155	9302	11,411	10,357	0.84	0.222	0.480	0.190	0.782	0.575	1935	2225
Pioneer 31D58	10,450	9770	12,119	10,945	0.911	0.266	0.772	0.490	0.926	0.633	1883	3198

VT3 had the largest difference at 12% and 15%, respectively, while for the other hybrids the differences were less than 5%.

3.4. Comparison between simulated and observed data

The combination of calibration and evaluation data presents a clear map for describing the performance of both crop models for all years and locations in simulating grain yield (Fig. 1). Because linear regression and related statistics could possibly mislead a performance analysis, in this study we also conducted a graphical analysis by comparing simulated with the observed data with reference to the 1:1 line. At first glance, many of the single simulations (year ∗ location) based on the model EPIC were higher than the observed yield, especially for the hybrids Pioneer 33M57(Hx1/LL.RR2), Croplan Genetics 8756 VT3, and SS731CL. In contrast to EPIC, the simulated yield for CSM-CERES-Maize was closer to the 1:1 line, especially for the hybrids Dyna-Gro V5373VT3, Pioneer 33M57 (Hx1/LL/RR2), and Pioneer 31D58, which means that the single simulations (year ∗ location) were fairly accurate. For the hybrid Dyna-Gro V5373VT3, EPIC tended to slightly overestimate for low observed grain yield values, while CSM-CERES-Maize showed more accurate simulations when the observed grain yield was lower. For the hybrid Pioneer 33M57 (Hx1/LL.RR2) andSS731CL, EPIC overestimated grain yield, while the CSM-CERES-Maize model showed that a scattered simulated yield for SS 731CL when compared to observed with a poor fit, but a good fit for the hybrid Pioneer 33M57 (Hx1/LL.RR2). For the hybrid Pioneer 31D58, both crop models showed

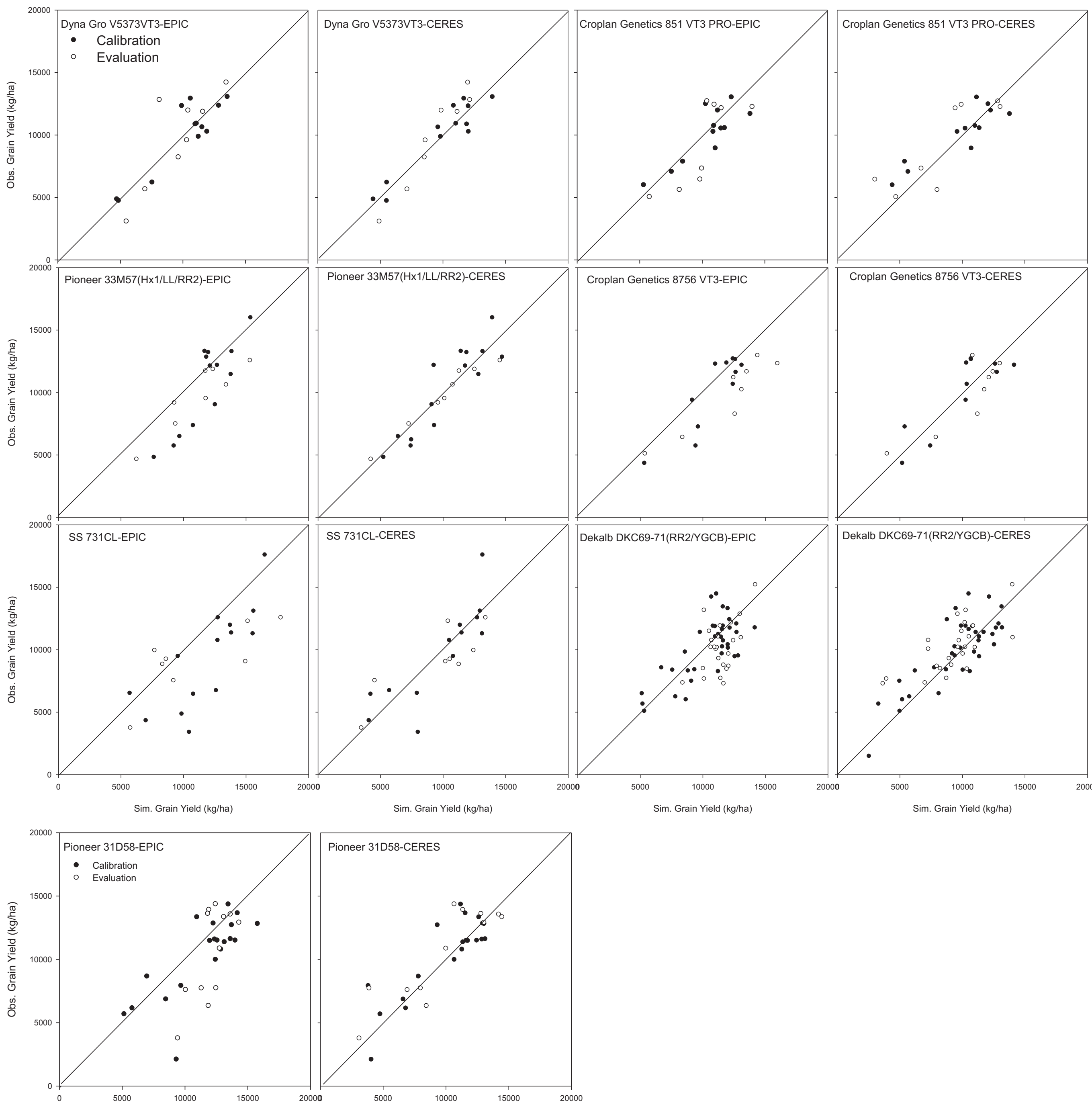

Fig. 1. A comparison between simulated and observed grain yield based on the CSM-CERES-Maize and EPIC models for calibration and evaluation of the seven hybrids and the 1:1 line.

a similar comparison with observed yield. For the hybrids Croplan Genetics 851 VT3 and Croplan Genetics 8756 VT3, both models provided accurate simulations when compared with the observed data. For the hybrid DeKalb DKC69-71(RR2/YGCB), EPIC tended to overestimate for the evaluation data set, which CSM-CERES-Maize tended to underestimate. In summary, the CSM-CERES-Maize showed a slightly better simulation of grain yield than EPIC especially for the hybrids SS731CL and Pioneer 31D58, while the two models were comparable in predicting grain yield for the other hybrids.

3.5. Comparison of long-term simulations

A long-term simulation analysis was conducted using 55 years of historical weather data, with the same crop management that was used for the variety trial data. For rainfed conditions the simulated grain yield for 55 years is summarized for both models in Fig. 2. The simulated grain yield for CSM-CERES-Maize ranged from 1000 kg/ha to 14,000 kg/ha, with a median yield ranging from 5500 kg/ha to 6500 kg/ha for the hybrid Dyna-Gro V5373VT3 at the six locations. A large range, e.g., the difference between the minimum and the maximum value was found among years due to the differences in precipitation for each year. Simulations with EPIC for the hybrid Dyna-Gro V5373VT3 were similar to CSM-CERES-Maize for Blairsville, but the minimum and maximum values were about 1000 kg/ha less. For Calhoun, the minimum, median, and maximum values for the simulations based on EPIC were about 3000 kg/ha higher than for CSM-CERES-Maize, while the yield simulations for EPIC for Griffin were similar to Blairsville. Although a similar median was found for both models at Midville, EPIC showed a smaller range. For Plains, the simulations based on EPIC had a maximum value of about 8200 kg/ha, which was much lower than for CSM-CERES-Maize. However, the yield predictions for both models had a similar median, and EPIC showed that about 50% of the simulations ranged from 6000 kg/ha to 7000 kg/ha. For Tifton, the median simulations based on EPIC were about 2000 kg/ha lower than for CSM-CERES-Maize, while the minimum values were about 2000 kg/ha higher. However, about 50% of simulations for EPIC ranged from 5000 kg/ha to 6,000 kg/ha, which was similar to Plains. The simulated yield for the other six hybrids for both models were similar Dyna-Gro V5373VT3 and are, therefore, not discussed in detail (Fig. 2).

For irrigated conditions the simulated yield for both models was much higher compared to the rainfed conditions and the range was much smaller, mainly because there was no water deficit and the variability of local rainfall was not an issue when compared to the rainfed conditions (Fig. 3). The irrigated grain yield based on CSM-CERES-

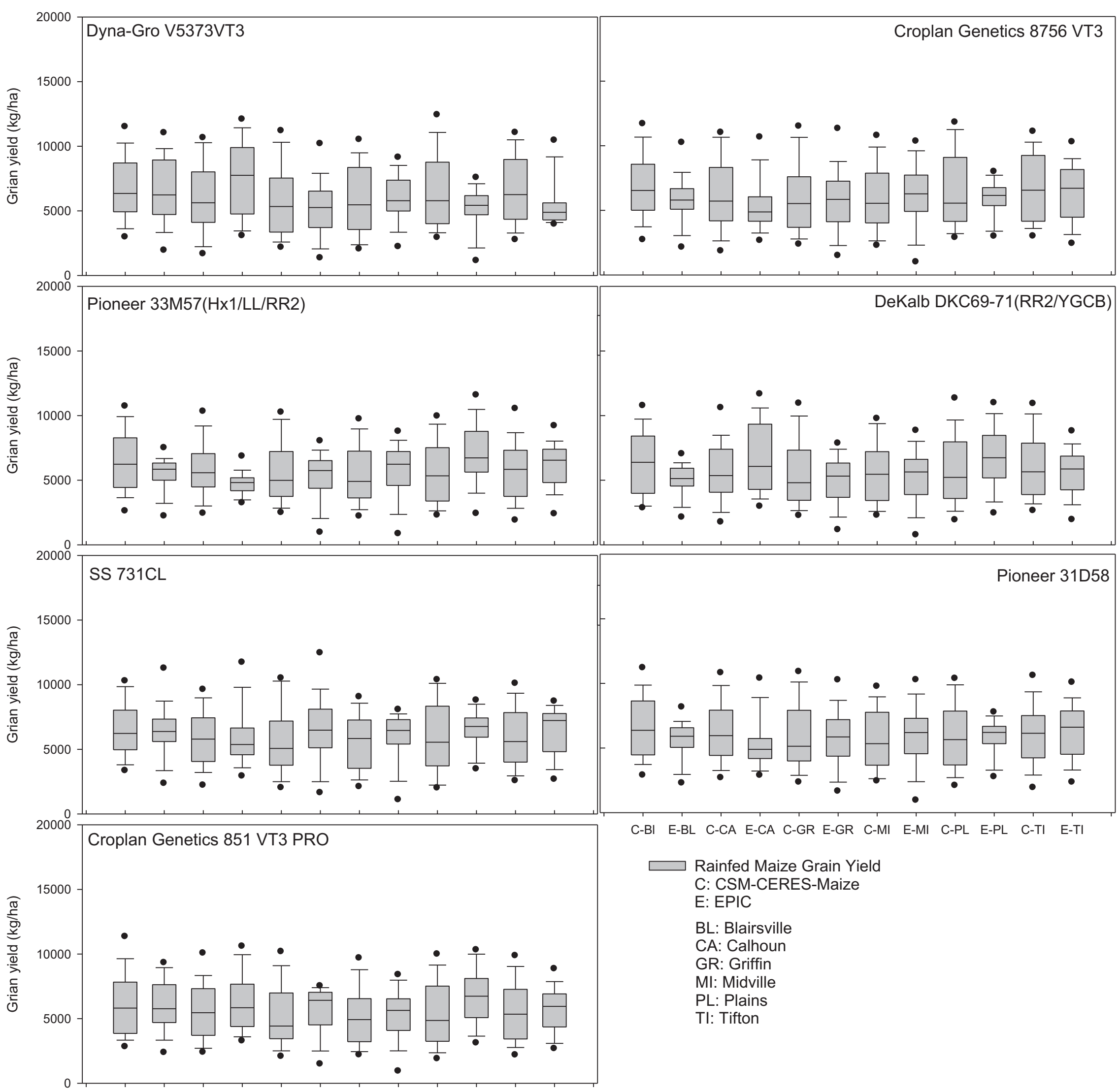

Fig. 2. Box-plot for rainfed grain yield based on the CSM-CERES-Maize and EPIC for seven hybrids using historical weather data from1958 to 2012 for Blairsville, Calhoun, Griffin, Midville, Plains, and Tifton, Georgia.

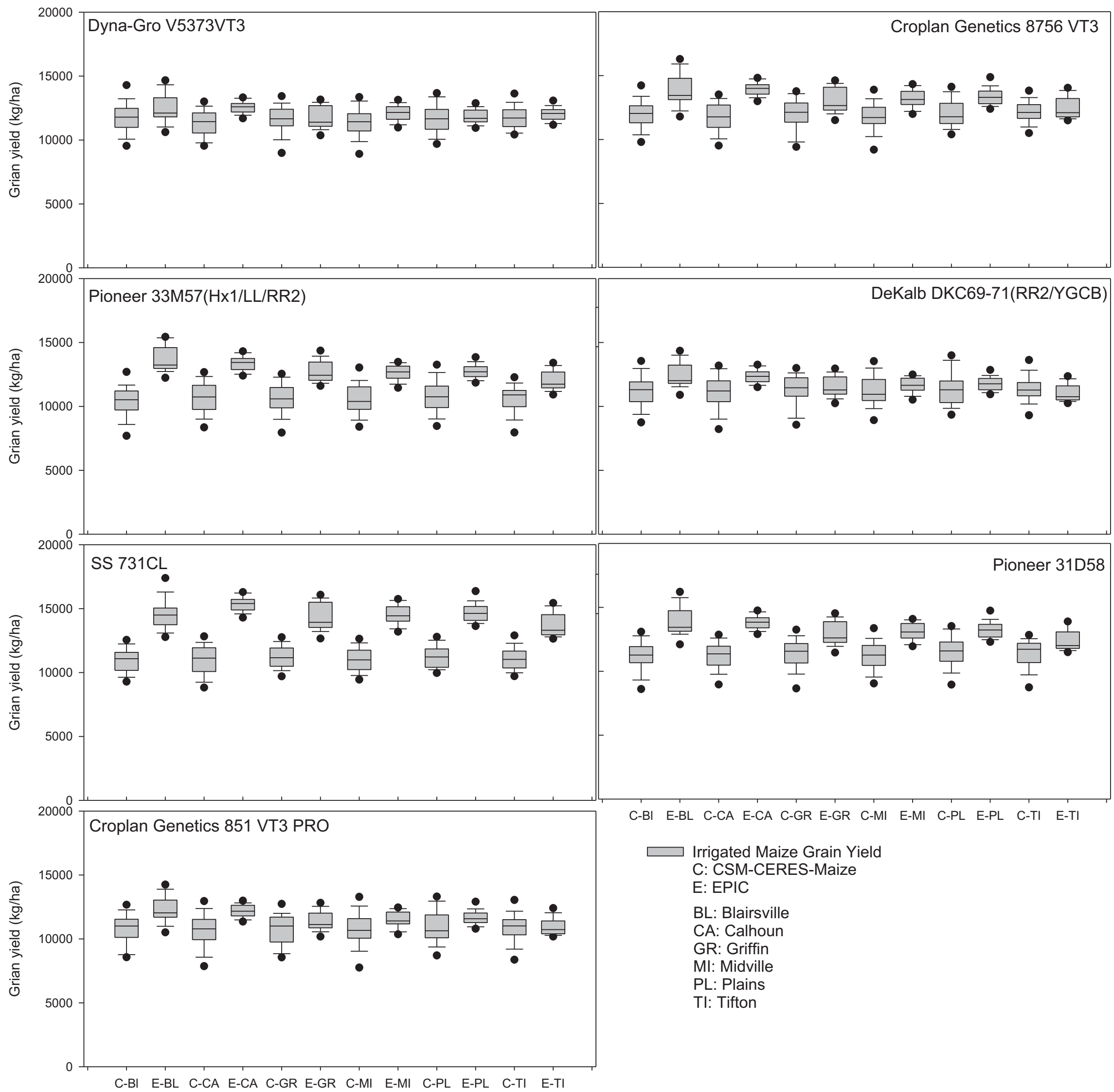

Fig. 3. Box-plot for irrigated grain yield based on CSM-CERES-Maize and EPIC for seven hybrids using historical weather data from 1958 to 2012 for Blairsville, Calhoun, Griffin, Midville, Plains, and Tifton, Georgia.

Maize ranged from about 8000 kg/ha to 15,000 kg/ha and the median was about 11,000 kg/ha for Dyna-Gro V5373VT3. The simulations based on EPIC had a very similar range when compared to CSM-CERES-Maize, but with a different median of around 12,000 kg/ha. Simulations for EPIC for Blairsville were higher than for the other locations. For both CSM-CERES-Maize and EPIC, the irrigated simulations with the hybrids Croplan Genetics 8756VT3, DeKalb DKC69-71 (RR2/YGCB), and Croplan Genetics 851VT3 PRO had a similar distribution compared to Dyn-Gro V5373VT3. However, for the hybrids Pioneer 33M57 (Hx1/LL/RR2), SS 731CL, and Pioneer 31D58 the differences were much larger, which was consistent with the earlier results found during calibration and evaluation (as shown in Fig. 1). In general, the simulations of Pioneer 33M57 (Hx1/LL/RR2), SS 731CL, and Pioneer 31D58 were very similar, ranging from 6500 kg/ha to 14,500 kg/ha for CSM-CERES-Maize and from 9000 kg/ha to 16,500 kg/ha for EPIC. However, simulations based on EPIC for Blairsville ranged from 11,000 kg/ha to 18,000 kg/ha for SS 731CL. The medians ranged from 11,000 kg/ha to 12,000 kg/ha for CSM-CERES-Maize and from 13,000 kg/ha to 14,000 kg/ha for EPIC.

The ANOVA test showed that the two crop models were significantly different for rainfed conditions for the hybrid Dyna-Gro V5373VT3 for Griffin, Plains, and Tifton; the hybrid Pioneer 33M57 (Hx1/LL/RR2) was significantly different for rainfed conditions for Griffin and Plains; the hybrid SS 731CL was significantly different for rainfed conditions for Blairsville; the hybrid Croplan Genetics 851 VT3 PRO was significantly different for rainfed conditions for Blairsville and Plains; the hybrid Croplan Genetics 8756 VT3 was significantly different for rainfed conditions for Blairsville and Calhoun; the hybrid DeKalb DKC69-71(RR2/YGCB) was significantly different for rainfed conditions for Calhoun and Griffin; the hybrid Pioneer 31D58 was significantly different for rainfed conditions for Calhoun. For irrigated conditions, the hybrid Dyna-Gro V5373VT3 was significantly different for Blairsville, Calhoun, and Midville; the hybrid Pioneer 33M57(Hx1/LL/RR2) and SS 731CL were significantly different for all locations, and for the hybrid Croplan Genetics 8756 VT3 the models were significantly different for Blairsville.

4. Discussion

This study conducted a calibration and evaluation for two commonly used maize crop models, CSM-CERES-Maize and EPIC, based on only observed grain yield for multiple years and locations in Georgia. Similar to prior studies, it was concluded that the CSM-CERES-Maize model can accurately simulate grain yield for different environments (Jagtap et al., 1993; Ritchie and Alagarswamy, 2003; Soler et al., 2007). The differences between simulated and observed yield was not more than 3% for calibration and not more than 8% for evaluation based on CSM-CERES-Maize, which means the simulations showed a good match with the observations. The statistical criteria, including slope, R^2, and RMSE, also showed a good fit, except R^2 for the hybrid DeKalb DKC69-71 (RR2/YGCB) had a value of 0.48, which was low.

Simulated grain yield was generally over-estimated by EPIC for all hybrids, with the differences between simulated and observed yield ranging from 2% to 23% for calibration and from 10 to 20% for evaluation, which were larger than for CSM-CERES-Maize. The EPIC simulations in this study were similar to those of Balkovič et al. (2013), who showed that EPIC underestimated for high yield conditions and overestimated for low yield conditions.

As discussed in many previous studies, all crop models suffer from considerable structural and parameter uncertainty and from a lack of independent datasets for thorough model evaluation (Knutti, 2010; Rötter et al., 2012). Prior to any model applications, it is important to demonstrate the confidence in predicting crop grain yield (Asseng et al., 2013; Carter, 2013). In this study, the performance of two maize simulation models was, in general, consistent and comparable for all seven hybrids that were evaluated. Both models provided the most accurate simulations for Dyna-Gro V5373VT3, Croplan Genetics 851 VT3, Pioneer 33M57(Hx1/LL.RR2), and Croplan Genetics 8756 VT3, and with less confidence for the hybrid DeKalb DKC69-71(RR2/YGCB). However, differences existed between the two crop models in simulating maize yield, which was caused by the differences in model structure and external parameters. For example, CSM-CERES-Maize showed more accurate simulations for the hybrids SS731CL and Pioneer 31D58. The combined simulations of CSM-CERES-Maize and EPIC were, in some cases, better than the single model simulations, but cannot necessarily be considered an ensemble.

In order to compare the response of the two models for long-term simulations, the same locations were used, but with 55 years of historical weather data and the same management as the variety trial data, the same hybrids, but for both rainfed and irrigated conditions. Applying irrigation eliminates the impact of rainfall variability and thus the potential impact of the water balance on the long-term model comparisons. Thus, the differences in rainfed grain yield were mainly caused due to differences in rainfall among years and locations, and due to the differences between the two crop model responses. For irrigated conditions, simulated grain yield was much higher compared to the rainfed yield, and showed a much smaller range in grain yield between the minimum and maximum values. The difference in grain yield among locations was not that significant, although for each location, the median yield for EPIC was higher than for CSM-CERES-Maize. Overall, the simulated rainfed and irrigated grain yield based on the two crop models was reasonable when compared to the earlier observations that were used for model calibration and evaluation.

5. Conclusion

The results from this study showed that long-term variety trial data that only include grain yield and final harvest dates can be used for the calibration of crop simulation models. The evaluation of the CSM-CERES-Maize and EPIC models with the observed independent data was accurate given the uncertainty of the observations. However, the long-term simulations with the two crop simulation models showed differences between the two models for some locations, which could potentially impact climate change and related application studies.

Acknowledgements

This research was supported in part by research grants associated with the Southeast Climate Consortium with funding received from NOAA-RISA (NA100AR4310215) and USDA-NIFA (2011-67003-30347).

References

Adams, R.M., Hurd, B.H., Lenhart, S., Leary, N., 1998. Effects of global climate change on agriculture: An interpretative review. Clim. Res. 11, 19–30.
Alexandrov, V.A., Hoogenboom, G., 2000. Vulnerability and adaptation assessments of agricultural crops under climate change in the Southeastern USA. Theor. Appl. Climatol. 67, 45–63.
Anothai, J., Patanothai, A., Jogloy, S., Pannangpetch, K., Boote, K., Hoogenboom, G., 2008. A sequential approach for determining the cultivar coefficients of peanut lines using end-of-season data of crop performance trials. Field Crop Res. 108, 169–178.
Asseng, S., Ewert, F., Rosenzweig, C., Jones, J.W., Hatfield, J.L., Ruane, A.C., Boote, K.J., Thorburn, P.J., Rotter, R.P., Cammarano, D., Brisson, N., Basso, B., Martre, P., Aggarwal, P.K., Angulo, C., Bertuzzi, P., Biernath, C., Challinor, A.J., Doltra, J., Gayler, S., Goldberg, R., Grant, R., Heng, L., Hooker, J., Hunt, L.A., Ingwersen, J., Izaurralde, R.C., Kersebaum, K.C., Muller, C., Naresh Kumar, S., Nendel, C., O/'Leary, G., Olesen, J.E., Osborne, T.M., Palosuo, T., Priesack, E., Ripoche, D., Semenov, M.A., Shcherbak, I., Steduto, P., Stockle, C., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Travasso, M., Waha, K., Wallach, D., White, J.W., Williams, J.R., Wolf, J., 2013. Uncertainty in simulating wheat yields under climate change. Nat. Clim. Chang. 3, 827–832.
Balkovič, J., van der Velde, M., Schmid, E., Skalský, R., Khabarov, N., Obersteiner, M., Stürmer, B., Xiong, W., 2013. Pan-European crop modelling with EPIC: Implementation, up-scaling and regional crop yield validation. Agric. Syst. 120, 61–75.
Bao, Y., Hoogenboom, G., McClendon, R.W., Paz, J., 2015. Potential adaptation strategies for rainfed soybean production in the south-eastern USA under climate change based on the CSM-CROPGRO-Soybean model. J. Agric. Sci. 153, 1–27.
Bassu, S., Brisson, N., Durand, J., Boote, K.J., Lizaso, J., Jones, J.W., Rosenzweig, C., Ruane, A.C., Adam, M., Baron, C.S., Basso, B., Biernath, C., Boogaard, H., Conijn, S., Corbeels, M., Deryng, D., De Sanctis, G., Gayler, S., Grassi, P., Hatfield, J., Hoek, S., Izaurralde, C., Jongschaap, R., Kemanian, A.R., Kersebaum, K.C., Kim, S., Kumar, N.S., Makowski, D., Muller, C., Nendel, C., Priesack, E., Pravia, M.V., Sau, F., Shcherbak, I., Tao, F., Teixeira, E., Timlin, D., Waha, K., 2014. How do various maize crop models vary in their responses to climate change factors? Glob. Chang. Biol. 20, 2301–2320.
Cabelguenne, M., Jones, C.A., Marty, J.R., Dyke, P.K., Williams, J.R., 1990. Calibration and validation of EPIC for crop rotation in Souhtern France. Agric. Syst. 33, 153–171.
Carbone, G.J., Kiechle, W., Locke, C., Mearns, L.O., McDaniel, L., Downton, M.W., 2003. Response of Soybean and Sorghum to varying spatial scales of climate change scenarios in the southeastern United States. Clim. Chang. 60, 73–98.
Carter, T.R., 2013. Agricultural impacts: multi-model yield projections. Nat. Clim. Chang. 3, 784–786.
Casella, G., Berger, R.L., 2002. Statistical Inference. Thomson Learning.
Ceglar, A., Črepinšek, Z., Kajfež-Bogataj, L., Pogačar, T., 2011. The simulation of phenological development in dynamic crop model: The Bayesian comparison of different methods. Agric. For. Meteorol. 151, 101–115.
Cerrato, M.E., Blackmer, A.M., 1990. Comparison of models for describing; corn yield response to nitrogen fertilizer. Agron. J. 82, 138–143.
Chai, T., Draxler, R.R., 2014. Root mean square error (RMSE) or mean absolute error (MAE)? – arguments against avoiding RMSE in the literature. Geosci. Model Dev. 7, 1247–1250.
Coy, A.E., Day, J.L., Rose, P.A., 2003. Georgia 2003 corn performance tests. Research Report Number 690, The Georgia Agricultural Experiment Stations, College of Agricultural and Environmental Sciences. The University of Georgia, Athens, GA.
Coy, A.E., Day, J.L., Rose, P.A., 2004. Georgia 2004 corn performance tests. Research Report Number 696, The Georgia Agricultural Experiment Stations, College of Agricultural and Environmental Sciences. The University of Georgia, Athens, GA.
Coy, A.E., Day, J.L., Rose, P.A., 2005. Georgia 2005 corn performance tests. Research Report Number 701, The Georgia Agricultural Experiment Stations, College of Agricultural and Environmental Sciences. The University of Georgia, Athens, GA.
Coy, A.E., Day, J.L., Rose, P.A., 2006. Georgia 2006 corn performance tests. Research Report Number 707, The Georgia Agricultural Experiment Stations, College of Agricultural and Environmental Sciences. The University of Georgia, Athens, GA.
Coy, A.E., Day, J.L., Rose, P.A., 2007. Georgia 2007 corn performance tests. Research Report Number 712, The Georgia Agricultural Experiment Stations, College of Agricultural and Environmental Sciences. The University of Georgia, Athens, GA.
Coy, A.E., Day, J.L., Rose, P.A., 2008. Georgia 2008 corn performance tests. Research Report Number 717, The Georgia Agricultural Experiment Stations, College of Agricultural and Environmental Sciences. The University of Georgia, Athens, GA.
Coy, A.E., Day, J.L., Rose, P.A., 2009. Georgia 2009 corn performance tests. Annual Publication Number 101, The Georgia Agricultural Experiment Stations, College of Agricultural and Environmental Sciences. The University of Georgia, Athens, GA.
Coy, A.E., Day, J.L., Rose, P.A., 2010. Georgia 2010 corn performance tests. Annual Publication Number 101–2, The Georgia Agricultural Experiment Stations, College of Agricultural and Environmental Sciences. The University of Georgia, Athens, GA.
Curry, R.B., Jones, J.W., Boote, K.J., Peart, R.M., Allen, L.H., Pickering, N.B., 1995. Response of soybean to predicted climate change in the USA. In: Rosenzweig, C., Ritchie, J.T., Jones, J.W., Tsuji, G.Y., Hildebrand, P., Allen Jr., L.H., Harper, L.A., Hollonger, S.E., Peterson, G.A., Kral, D.M., MK, V. (Eds.), Climate Change and Agriculture: Analysis of Potential International Impacts. ASA Special Publication Number. 59, pp. 163–182.
Easterling, W.E., Chen, X., Hays, C.J., Brandle, J.R., Zhang, H., 1996. Improving the validation of model-simulated crop yield response to climate change: an application to the EPIC model. Clim. Res. 6, 263–273.
Easterling, W.E., Hays, C.J., Easterling, M.M., Brandle, J.R., 1997. Modelling the effect of shelterbelts on maize productivity under climate change: an application of the EPIC model. Agric. Ecosyst. Environ. 61, 163–176.
Fleisher, D.H., Condori, B., Quiroz, R., Alva, A., Asseng, S., Barreda, C., Bindi, M., Boote, K.J., Ferrise, R., Franke, A.C., Govindakrishnan, P.M., Harahagazwe, D., Hoogenboom, G., Naresh Kumar, S., Merante, P., Nendel, C., Olesen, J.E., Parker, P.S., Raes, D.,

Raymundo, R., Ruane, A.C., Stockle, C., Supit, I., Vanuytrecht, E., Wolf, J., Woli, P., 2016. A Potato model inter-comparison across varying climates and productivity levels. Glob. Chang. Biol. (Accepted for publication).
French, V., Hodges, T., 1985. Comparison of crop phenology models. Agron. J. 77, 170–171.
Gaiser, T., de Barros, I., Sereke, F., Lange, F.-M., 2010. Validation and reliability of the EPIC model to simulate maize production in small-holder farming systems in tropical sub-humid West Africa and semi-arid Brazil. Agric. Ecosyst. Environ. 135, 318–327.
Garcia y Garcia, A., Hoogenboom, G., 2005. Evaluation of an improved daily solar radiation generator for the southeastern USA. Clim. Res. 29, 91–102.
Guerra, L.C., Hoogenboom, G., Boken, V.K., Hook, J.E., Thomas, D.L., Harrison, K.A., 2004. EPIC model for simulating crop yield and irrigation demand. Trans. ASABE 47, 2091–2100.
Guerra, L.C., Hoogenboom, G., Garcia y Garcia, A., Banterng, P., Beasley, J.P., 2008. Determination of cultivar coefficients for the CSM-CROPGRO-Peanut model using variety trial data. Trans. ASABE 54, 1471–1481.
Holzworth, D.P., Snow, V., Janssen, S., Athanasiadis, I.N., Donatelli, M., Hoogenboom, G., White, J.W., Thorburn, P., 2015. Agricultural production systems modelling and software: current status and future prospects. Environ. Model. Softw. 72 (1), 276–286.
Hoogenboom, G., 1996. The Georgia automated environmental monitoring network. 22th Agricultural and Forest Meteorology Conference. American Meteorological Society, Atlanta, Georgia, pp. 343–346.
Hoogenboom, G., 2000. Contribution of agrometeorology to the simulation of crop production and its applications. Agric. For. Meteorol. 103, 137–157.
Hoogenboom, G., Jones, J.W., Traore, P.C.S., Boote, K.J., 2012a. Experiments and data for model evaluation and application. In: Kihara, J., D.F., Jones, J.W., Hoogenboom, G., Tabo, R., Bationo, A. (Eds.), Improving Soil Fertility Recommendations in Africa using the Decision Support Systems for Agrotechnology Transfers (DSSAT). Springer, Dordrecht, the Netherlands, pp. 9–18.
Hoogenboom, G., Jones, J.W., Wilkens, P.W., Porter, C.H., Boote, K.J., Hunt, L.A., Singh, U., Lizaso, J.L., White, J.W., Uryasev, O., Royce, F.S., Ogoshi, R., Gijsman, A.J., Tsuji, G.Y., Koo, J., 2012b. Decision Support System for Agrotechnology Transfer (DSSAT) Version 4.5 [CD-ROM]. University of Hawaii, Honolulu, Hawaii.
Hunt, L.A., Boote, K.J., 1998. Data for model operation, calibration, and evaluation. In: Tsuji, G.Y., Hoogenboom, G., Thornton, P.K. (Eds.), Understanding Options for Agricultural Production. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 9–39.
Hunt, L.A., Pararajasingham, S., Jones, J.W., Hoogenboom, G., Imamura, D.T., Ogoshi, R.M., 1993. GENCALC: software to facilitate the use of crop models for analyzing field experiments. Agron. J. 85, 1090–1094.
Hunt, L.A., White, J.W., Hoogenboom, G., 2001. Agronomic data: advances in documentation and protocols for exchange and use. Agric. Syst. 70, 477–492.
Jagtap, S.S., Mornu, M., Kang, B.T., 1993. Simulations of growth, development, and yield of maize in the transition zone of Nigeria. Agric. Syst. 41, 215–229.
Jones, C.A., Kiniry, J.R., 1986. CERES-Maize: a Simulation Model of Maize Growth and Development. Texas A&M University Press, College Station, TX.
Jones, C.A., Sharpley, A.N., Williams, J.R., 1984a. A simplified soil and plant phosphorus model: I. Documentation. Soil Sci. Soc. Am. J. 48, 800–805.
Jones, C.A., Sharpley, A.N., Williams, J.R., 1984b. A simplified soil and plant phosphorus model: III. Testing. Soil Sci. Soc. Amer. J. 48, 800–805.
Jones, J.W., Boote, K.J., Hoogenboom, G., Jagtap, S.S., Wilkerson, G.G., 1989. SOYGRO V5.42, Soybean crop growth simulation model. User's guide. Florida Agric Exp. St. J. 8304. International Benchmark Sites Network for Agrotechnology Transfer (IBSNAT) and University of Florida, Gainesville, FL.
Jones, J.W., Hoogenboom, G., Porter, C.H., Boote, K.J., Batchelor, W.D., Hunt, L.A., Wilkens, P.W., Singh, U., Gijsman, A.J., Ritchie, J.T., 2003. The DSSAT cropping system model. Eur. J. Agron. 18, 235–265.
Kersebaum, K.C., Boote, K.J., Jorgenson, J.S., Nendel, C., Bindi, M., Frühauf, C., Gaiser, T., Hoogenboom, G., Kollas, C., Olesen, J.E., Rötter, R.P., Ruget, F., Thorburn, P., Trnka, M., Wegehenkel, M., 2015. Analysis and classification of data sets for calibration and validation of agro ecosystem models. Environ. Model. Softw. 72 (1), 402–417.
Knutti, R., 2010. The end of model democracy? Clim. Chang. 102, 395–404.
Ko, J., Piccinni, G., Guo, W., Steglich, E., 2009. Parameterization of EPIC crop model for simulation of cotton growth in South Texas. J. Agric. Sci. 147, 169–178.
Liu, B., Asseng, S., Müller, C., Ewert, F., Elliott, J., Lobell, D., Martre, P., Ruane, A., Wallach, D., Jones, J.W., Rosenzweig, C., Aggarwal, P., Alderman, P., Anothai, J., Basso, B., Biernath, C., Cammarano, D., Challinor, A., Deryng, D., De Sanctis, G., Doltra, J., Fereres, E., Folberth, C., Garcia-Vila, M., Gayler, S., Hoogenboom, G., Hunt, L., Izaurralde, R., Jabloun, M., Jones, C., Kersebaum, K., Kimball, B., Koehler, A., Kumar, S.N., Nendel, C., O'Leary, G., Olesen, J., Ottman, M., Palosuo, T., Prasad, P., Priesack, E., Pugh, T., Reynolds, M., Rezaei, E., Rötter, R.P., Schmid, E., Semenov, M., Shcherbak, I., Stehfest, E., Stöckle, C., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Thorburn, P., Waha, K., Wall, G., Wang, E., White, J.W., Wolf, J., Zhao, Z., Zhu, Y., 2016. Similar estimates of temperature impacts on global wheat yield by three independent methods. Nat. Clim. Chang. (In Press).
Lobell, D.B., Asner, G.P., 2003. Climate and management contributions to recent trends in U.S. agricultural yields. Science 299, 1032.
Mavromatis, T., Boote, K., Jones, J., Irmak, A., Shinde, D., Hoogenboom, G., 2001. Developing genetic coefficients for crop simulation models with data from crop performance trials. Crop Sci. 41, 40–51.
Parry, M., Rosenzweig, C., Iglesias, A., Livermore, M., Fischer, G., 2004. Effects of climate change on global food production under SRES emissions and socio-economic scenarios. Glob. Environ. Chang. 14, 53–67.
Parry, M.L., Canziani, O.F., Palutikof, J.P., van der Linden, P.J., Hanson, C.E., 2007. Climate change 2007: impacts, adaptation and vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, UK, p. 976.
Perez-Quezada, J.F., Cavero, J., Williams, J.R., Roel, A., Plant, R.E., 2003. Simulation of within-field yield variability in a four-crop rotation field using SSURGO soil-unit definitions and the EPIC model. Trans. ASABE 46, 1365–1374.
Perkins, H.F., Moss, R.B., Hutchins, A., 1978. Soils of the Southwest Georgia Stations. Research Bulletin 217. Georgia Agricultural Experiment Stations, Athens, GA.
Perkins, H.F., McCreery, R.A., Lockaby, G., Perry, C.E., 1979. Soils of the Southeast Georgia branch experiment station. Research Bulletin 245. the University of Georgia College of Agriculture Experiment Stations, Athens, GA.
Perkins, H.F., Owen, V., Hammel, J.E., Price, E.A., 1982. Soil characteristics of the plant science farm of the University of Georgia College Experiment Station. Research Bulletin 287. the University of Georgia College of Agriculture Experiment Stations, Athens, GA.
Perkins, H.F., Owen, V.R., Worley, E.E., 1983. Soils of the Northwest Georgia Research Experiment Station. Research Bulletin 302. College of Agriculture Experiment Stations. the University of Georgia, Athens, GA.
Perkins, H.F., Schuman, L.M., Boswell, F.C., Owen, V., 1985. Soil characteristics of the Bledsoe and Beckham research farms of the Georgia Station. Research Bulletin 332. the University of Georgia Agricultural Experiment Stations, Athens, GA.
Perkins, H.F., Hook, J.E., Barbour, N.W., 1986. Soil characteristics of selected areas of the Coastal Plain Experiment Station and ABAC Research Farms. Research Bulletin 346. The Georgia Agricultural Experiment Stations, College of Agriculture. the University of Georgia, Athens, GA.
Porter, J.R., Jamieson, P.D., Wilson, D.R., 1993. Comparison of the wheat simulation models Afrcwheat2, Ceres-wheat and Swheat for non-limiting conditions of crop growth. Field Crop Res. 33, 131–157.
Ritchie, J.T., Alagarswamy, G., 2003. Model concepts to express genetic differences in maize yield components. Agron. J. 95, 4–9.
Ritchie, J.T., Singh, U., Godwin, D.C., Bowen, W.T., 1998. Cereal growth, development and yield. In: Tsuji, G., Hoogenboom, G., Thornton, P. (Eds.), Understanding Options for Agricultural Production. Springer, Netherlands, pp. 79–98.
Rosenzweig, C., Jones, J.W., Hatfield, J.L., Ruane, A.C., Boote, K.J., Thorburn, P., Antle, J.M., Nelson, G.C., Porter, C., Janssen, S., Asseng, S., Basso, B., Ewert, F., Wallach, D., Baigorria, G., Winter, J.M., 2013. The Agricultural Model Intercomparison and Improvement Project (AgMIP): protocols and pilot studies. Agric. For. Meteorol. 170, 166–182.
Rötter, R.P., Palosuo, T., Kersebaum, K.C., Angulo, C., Bindi, M., Ewert, F., Ferrise, R., Hlavinka, P., Moriondo, M., Nendel, C., Olesen, J.E., Patil, R.H., Ruget, F., Takáč, J., Trnka, M., 2012. Simulation of spring barley yield in different climatic zones of Northern and Central Europe: a comparison of nine crop models. Field Crop Res. 133, 23–36.
Semenov, M.A., Shewry, P.R., 2011. Modelling predicts that heat stress, not drought, will increase vulnerability of wheat in Europe. Sci. Rep. 1.
Semenov, M.A., Stratonovitch, P., 2010. Use of multi-model ensembles from global climate models for assessment of climate change impacts. Clim. Res. 41, 1–14.
Sharpley, A.N., Jones, C.A., Gray, C., Cole, C.V., 1984. A simplified soil and plant phosphorus model: II. Prediction of labie, organic, and sorbed phosphorus. Soil Sci. Soc. Amer. J. 48, 800–805.
Soler, C., Sentelhas, P., Hoogenboom, G., 2007. Application of the CSM-CERES-Maize model for planting date evaluation and yield forecasting for maize grown off-season in a subtropical environment. Eur. J. Agron. 27, 165–177.
White, J.W., Hoogenboom, G., 2003. Gene-based approaches to crop simulation: past experiences and future opportunities. Agron. J. 95 (1), 52–64.
White, J.W., Hoogenboom, G., 2010. Crop response to climate: Ecophysiological models. In: Lobell, D., Burke, M. (Eds.), Climate Change and Food Security: Adapting Agriculture to a Warmer World. Springer, Dordrecht Herdelberg London New York, pp. 59–83.
White, J.W., Hoogenboom, G., Kimball, B.A., Wall, G.W., 2011. Methodologies for simulating impacts of climate change on crop production. Field Crop Res. 124, 357–368.
White, J.W., Hunt, L.A., Boote, K.J., Jones, J.W., Koo, J., Kim, S., Porter, C.H., Wilkens, P.W., Hoogenboom, G., 2013. Integrated description of agricultural field experiments and production: The ICASA Version 2.0 Data Standards. Comput. Electron. Agric. 96 (1), 1–12.
Williams, J.R., Jones, C.A., Dyke, P.K., 1984. A modeling approach to determining the relationship between erosion and soil productivity. Trans. ASABE 27, 129–144.
Williams, J.R., Jones, C.A., Kiniry, J.R., Spaniel, D.A., 1989. The EPIC Growth Model. Trans. ASABE 32, 479–511.
Wolf, J., 2002. Comparison of two soya bean simulation models under climate change. I. Model calibration and sensitivity analyses. Clim. Res. 20, 55–70.
Yang, J.M., Yang, J.Y., Liu, S., Hoogenboom, G., 2014a. An evaluation of the statistical methods for testing the performance of crop models with observed data. Agric. Syst. 127, 81–89.
Yang, J.Y., Drury, C.F., Yang, J.M., Li, Z.T., Hoogenboom, G., 2014b. EasyGrapher: software for data visualization and statistical evaluation of DSSAT cropping system model and the CANB model. International Journal of Computer Theory and Engineering 6, 210–214.