

# AGRONOMIC MODELING

## Modifying the CROPGRO-Soybean Model to Improve Predictions for the Upper Midwest

P. Pedersen,\* K. J. Boote, J. W. Jones, and J. G. Lauer

### ABSTRACT

The CROPGRO-Soybean model has not been extensively evaluated in the upper Midwest. The objective of this project was to determine if modifications of the CROPGRO-Soybean model would improve predictions in the upper Midwest using three cultivars in five management systems and two planting dates from 1997 to 2000. Version 3.5 of the model was compared with 1998 data and found to underestimate total biomass and grain yield at harvest. Changes in temperature function on leaf expansion rate and base temperature for pod addition improved model performance and decreased root mean square error (RMSE) for biomass at harvest and grain yield from 734 to 707 kg ha<sup>-1</sup> and from 410 to 362 kg ha<sup>-1</sup>, respectively. The modified model was then tested with independent data from 1997, 1999, and 2000. Overall, the model parameters calibrated from 1998 data improved the fit slightly but with higher RMSE values for the three independent years than the 1998 data set. Averaged across the 3 yr, the modified model underpredicted biomass at harvest and grain yield by 14 and 6%, respectively, with RMSE for biomass at harvest and grain yield averaging 1181 and 814 kg ha<sup>-1</sup>, respectively. The inaccuracy was related to underprediction of early vegetative growth because of the effect of site-specific and planting date-specific differences in temperature on biomass accumulation and leaf area index. It was concluded that the modified parameters improved the accuracy of the CROPGRO-Soybean model for the calibration year but did not significantly improve prediction for the three independent years.

THERE HAS BEEN an increased interest in modeling soybean [*Glycine max* (L.) Merr.] growth to predict vegetative and reproductive performance of different cultivars under various management systems and environmental conditions. However, predicting grain yield is difficult because of the wide ranges in compensatory effect on yield components, growth habit, and reproductive development of soybean cultivars (Cooper, 1977).

Crop models are available for almost all economically important crops and have the ability to predict yield and evaluate different options to maximize profit and/or minimize losses of nutrients or chemicals by integrating the effects of daily weather data with soil characteristics and management practices (Boote et al., 1998). Dynamic crop models have potential to quantify the contribution

of environmental factors, such as temperature, daylength, soil characteristics, and water supply. They are also suitable for studying variations in complex genotype X environment interactions (Hunt et al., 1993; Mavromatis et al., 2001).

In CROPGRO-Soybean, the effects of temperature on photosynthesis are calculated with leaf-level functions derived from the data of Harley et al. (1985) on light and CO<sub>2</sub>-saturated photosynthesis of soybean leaves in response to temperature. The derived function for CROPGRO-Soybean uses a linear function for photosynthetic electron transport with a base temperature of 8°C to an optimum of 40°C, with increasing leaf damage above 40°C that progressively reduces the rate to zero at 48°C (Boote et al., 1998).

Since soybean is chill sensitive and soil temperature affects soybean emergence, root development, and N transformations, low soil temperatures commonly observed during the early planting period or under conservation tillage practices in the upper Midwest will delay soybean emergence and adversely affect stand establishment with an underestimation of final yield. Allen et al. (1996), Andales et al. (2000), and Sexton et al. (1998) found that the CROPGRO-Soybean model did not predict soil temperature and emergence well under cool, wet conditions, which may translate to errors in timing of biomass accumulation during the remainder of the season.

Another potential problem is the prediction of soybean phenology. The accuracy of final yield predictions depends on timely predictions of critical growth stages beginning with emergence. However, predicting soybean phenology is difficult because of lack of understanding of sensitivity to temperature and photoperiod during development (Grimm et al., 1994), with sensitivity of soybean development rate to cool temperature decreasing after beginning seed fill (Grimm et al., 1994; Seddigh et al., 1989).

A major barrier to use of crop models is the lack of information required to run the models as well as the complexities of calibrating and validating them across different environments. A major reason contributing to criticisms of crop models is the scarcity of appropriate data for running and testing the models (Grant, 1989; O'Leary et al., 1985). Common practice has been to split available data into two groups: one for parameter estimation and the other for testing. However, with limited data, such splitting may result in less-accurate pa-

P. Pedersen, Dep. of Agron., Iowa State Univ., 2104 Agronomy Hall, Ames, IA 50011; K.J. Boote, Dep. of Agron., and J.W. Jones, Dep. of Agric. and Biol. Eng., Univ. of Florida, Gainesville, FL 32611; and J.G. Lauer, Dep. of Agron., Univ. of Wisconsin, 1575 Linden Dr.—Moore Hall, Madison, WI 53706. Received 12 May 2003. \*Corresponding author (palle@iastate.edu).

Published in *Agron. J.* 96:556–564 (2004).  
© American Society of Agronomy  
677 S. Segoe Rd., Madison, WI 53711 USA

**Abbreviations:** DAE, days after emergence; RMSE, root mean square error.

parameter estimates and prediction variances (Jones and Carberry, 1994).

Models are commonly developed based on rather limited field and controlled environmental data, and often they give unsatisfactory predictions when applied to field locations with different environmental conditions. The CROPGRO model was originally developed in a subtropical environment in Florida and tested primarily with Florida, Georgia, and North Carolina data sets, using only a few and limited studies from Ohio and Iowa under cooler environments (Boote et al., 1997). Other than the work of Sexton et al. (1998) and Sau et al. (1999), the CROPGRO-Soybean model has not been extensively evaluated under cooler environments. Our hypothesis is that the model, with existing relationships and parameters, underpredicts growth and yield during vegetative growth under cooler temperature. Furthermore, it was hypothesized that refinements in temperature functions in CROPGRO-Soybean could be made to facilitate model use over all weather conditions in Wisconsin. Therefore, the objective of this research was to determine if modifications to temperature functions in the CROPGRO-Soybean model that affect development and growth processes would improve predictions across environments in the upper Midwest. The model was evaluated using field data from 1998, and model species parameters were modified based on 1998 data to enable a more accurate prediction of crop growth and yield. The model with modified parameters was then validated against independent field data on the same three cultivars grown in different management systems and planting dates in 1997, 1999, and 2000.

## MATERIALS AND METHODS

### Field Data

Field experiments were conducted during 4 yr (1997–2000) in five different management systems. These management systems were chosen to represent current management practices in the upper Midwest. Four of the five management systems were conducted on a Plano silt loam soil (fine-silty, mixed, mesic, Typic Argiudoll) at the Arlington, WI, Agricultural Research Station (Fig. 1). They consisted of two conventional tillage and two no-tillage systems, both tillage systems with and without irrigation. The fifth management system was conducted at Hancock Agricultural Research Station on a Plainfield sandy loam soil (loamy-sand, mixed, mesic, Typic Udipsamment; Fig. 1). This management system was a conventional tillage system with irrigation. The experimental design for each management system was a randomized complete block in a split-plot arrangement with four replications. Main plot was planting date (early May vs. late May). The subplots were three soybean cultivars from maturity group II (Hardin, DeKalb CX232, and Spansoy 250). Cultivars were chosen to represent current vs. old cultivars in the upper Midwest. Management practices and descriptions of the management systems have previously been described in a companion paper and will therefore not be mentioned here (Pedersen and Lauer, 2003).

Sections of 0.76 m<sup>2</sup> were hand-harvested and used to determine dry matter accumulation and distribution on 21-d intervals starting 21 d after emergence (DAE). There were six sampling dates throughout the growing season (21, 42, 63, 84,

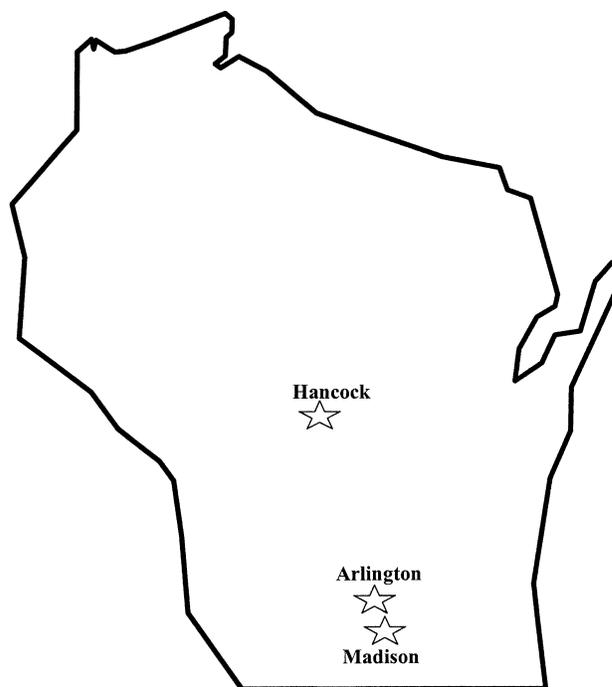


Fig. 1. Map showing the location of the Arlington and Hancock sites in Wisconsin.

105, and 126 DAE). Development and growth stage and plant height information were taken based on a sample of three plants randomly collected from the hand-harvested section and then separated into leaves, stems, pods, and seeds. Plant growth stages were determined according to Fehr and Caviness (1977).

### CROPGRO-Soybean Model

Soybean crop models have been available since the 1980s when the original version of SOYGRO V4.2 was released (Wilkerson et al., 1983). The model has substantially been modified from the original version, and CROPGRO-Soybean v. 3.5 is the most recent version (Boote et al., 1998; Hoogenboom et al., 1994). CROPGRO-Soybean is a process-oriented model that can be used to study soybean response to management (Egli and Bruening, 1992), environmental conditions (Curry et al., 1995), and genetic yield potential (Boote and Tollenaar, 1994). It also has been used to study causes of spatial yield variability (Allen et al., 1996; Paz et al., 1998).

The CROPGRO-Soybean model requires inputs of management practices and environmental conditions and incorporates knowledge of cultivar-specific traits (genetic coefficients) to predict daily growth and development as the plant responds to weather, soil characteristics, and management practices (Boote et al., 1998; Paz et al., 1998). The CROPGRO-Soybean model is generic and has a species and a cultivar data file. The species file describes species characteristics such as tissue composition and partitioning traits and includes sensitivity of processes to temperature, light, plant water deficit, and plant N deficiency. The cultivar data file includes information on life cycle phases, vegetative traits, leaf traits, potential seed fill duration, seed size, and seed composition (Boote et al., 2001; Hoogenboom et al., 1994; Jones, 1993; Jones et al., 1994).

### Model Modification and Evaluation Process

Water-holding traits for the two soils were determined by University of Wisconsin Soil and Plant Analysis Laboratory,

**Table 1. Genetic coefficients for the CROPGRO-Soybean model and definitions and units.**

Variables	Definition	Units
CSDL	Critical short daylength below which reproductive development progresses with no daylength effect	h
PPSEN	Slope of the relative response of development to photoperiod with time	h <sup>-1</sup>
EM-FL	Time between emergence and first flower	PTP†
FL-SH	Time between first flower and first pod	PTP
FL-SD	Time between first flower and first seed	PTP
FL-LF	Time between first flower and end of leaf expansion	PTP
SD-PM	Time between first seed and physiological maturity	PTP
LFMAX	Maximum leaf photosynthesis rate at 30°C, 350 vpm CO <sub>2</sub> , and high light	mg CO <sub>2</sub> m <sup>-2</sup> s <sup>-1</sup>
SLAVR	Specific leaf area of cultivar under standard growing conditions	cm <sup>2</sup> g <sup>-1</sup>
SIZLF	Maximum size of full leaf (three leaflets)	cm <sup>2</sup>
XFRT	Maximum fraction of daily growth that is partitioned to seed + shell	
SFDUR	Seed-filling duration for pod cohort at standard growth conditions	PTP
PODUR	Time required for cultivar to reach final pod load under optimal conditions	PTP
WTPSD	Maximum weight per seed	g
SDPDV	Average seeds per pod under standard growing conditions	no. pod <sup>-1</sup>

† Photothermal days.

Madison. The soil fertility factor (SLPF) was set to 1.0 for both locations, the default value for all Midwest soils by CROPGRO. SLPF is an input variable (constant for a given field site) that affects crop growth rate by modifying daily canopy photosynthesis. SLPF is attributed to soil fertility differences or soil-based pests, such as nematodes.

Most of the improvements made to the model parameters resulted from modifying the temperature function for pod addition and vegetative leaf expansion. The modification of the coefficients was initially based on previous work with soybean and common bean (Hume and Jackson, 1981; Marowitch et al., 1986; Caufield and Bunce, 1988; Sexton et al., 1994; Piper et al., 1996). In addition, temperature function before flowering for vegetative development was changed in the species file from linear to a sinus curve.

The 1998 experimental data set was used to calibrate the model while the 1997, 1999, and 2000 data were reserved for validating the modified model. The 1998 experimental data set was selected to calibrate the model since it was the first year where all management systems were present. The GENCALC program (Hunt et al., 1993) was used to calibrate the genetic coefficients (Table 1) of Hardin, CX232, and Spansoy 250 from the 1998 experimental data (Table 2). These coefficients were used for the 3 yr of validation.

The criteria by Wallach and Goffinet (1987) were used to assess the degree of improvement in model predictability: a decrease in RMSE but also improved intercept (*a*) and slope (*b*) values of linear regression between predicted and observed data and a higher index of agreement (*d*; Willmott, 1982). Root mean square error reflects the magnitude of the mean

difference between predicted and observed values over time and is calculated from the following equation:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i^o - y_i^p)^2}{n}} \quad [1]$$

where *n* is the number of observations, *y<sub>i</sub><sup>o</sup>* is the observed variable, and *y<sub>i</sub><sup>p</sup>* is the predicted variable for the specific cultivar. Index of agreement was chosen instead of Pearson's product-moment correlation coefficient, which can be misleading (Willmott and Wicks, 1980). Willmott and Wicks (1980) observed that the product-moment correlation coefficient often is unrelated to the sizes of the differences between the observed and the predicted observations. Index of agreement is calculated from the following equation:

$$d = 1 - \left[ \frac{\sum_{i=1}^N (P_i - O_i)^2}{\sum_{i=1}^N (|P_i| + |O_i|)^2} \right], 0 \leq d \leq 1 \quad [2]$$

where *P<sub>i</sub>* = predicted variable, *O<sub>i</sub>* = observed variable, *P<sub>i</sub>* = *P<sub>i</sub>* -  $\bar{O}$ , and *O<sub>i</sub>* = *O<sub>i</sub>* -  $\bar{O}$ .

## RESULTS AND DISCUSSION

Weather patterns are dominant factors controlling yield and soybean development in the upper Midwest, and the four growing seasons for this study produced unique effects on plant growth and development from which to evaluate weather effects (Pedersen and Lauer, 2003). Since rainfall was slightly above normal during all years, none of the treatments in any year or at any soil depth approached the permanent wilting point of 0.10 kg kg<sup>-1</sup> for a silt loam soil and 0.05 kg kg<sup>-1</sup> for a sandy loam soil (Schulte and Walsh, 1994). Thus, there was only a small effect of irrigation on the silt loam soil. Temperature and solar radiation are presented in Tables 3 and 4 for Arlington and Hancock, respectively. Small yield differences were observed for individual years among cultivars, planting dates, and management systems (Pedersen and Lauer, 2003) due to large compensatory responses for yield components as a result of considerable changes in growth and development (Pedersen, 2002). This discussion will mainly focus on bio-

**Table 2. Genetic coefficients of three soybean cultivars for the CROPGRO-Soybean Model as estimated in this study.**

Genetic coefficients	Default values†	Hardin	CX232	Spansoy 250
CSDL, h	13.59	13.60	13.36	13.55
PPSEN, h <sup>-1</sup>	0.249	0.249	0.249	0.249
EM-FL, ph. days‡	17.40	12.30	12.00	12.00
FL-SH, ph. days	6.0	6.00	6.00	7.00
FL-SD, ph. days	13.5	9.00	9.00	9.50
SD-PM, ph. days	33.00	44.05	41.44	47.30
FL-LF, ph. days	26.00	37.66	37.91	44.55
LFMAX, mg CO <sub>2</sub> m <sup>-2</sup> s <sup>-1</sup>	1.03	1.036	1.100	1.097
SLAVR	375	382	388	364
SIZLF	180	180	180	180
XFRT	1.0	1.0	1.0	1.0
WTPSD	0.1900	0.1053	0.1504	0.1180
SFDUR	23.00	21.12	26.99	25.83
SDPDV	2.200	2.342	2.513	2.650
PODUR	10.00	9.00	10.50	19.00

† Default values for maturity group 2.

‡ Photothermal days.

**Table 3. Mean daily solar radiation (SR) and mean daily maximum ( $T_{max}$ ) and minimum ( $T_{min}$ ) temperatures between April and September at Arlington from 1997 to 2000.**

Period	1997			1998			1999			2000		
	SR	$T_{max}$	$T_{min}$									
	MJ m <sup>-2</sup>	°C		MJ m <sup>-2</sup>	°C		MJ m <sup>-2</sup>	°C		MJ m <sup>-2</sup>	°C	
1–10 Apr.	20.1	11.5	-1.1	14.9	11.1	1.6	12.3	10.3	1.2	19.9	11.0	-1.6
11–20 Apr.	17.0	8.8	-2.5	18.6	15.2	3.0	24.0	15.3	2.2	11.5	10.7	1.3
21–30 Apr.	22.8	16.3	1.0	21.3	17.6	3.1	21.5	20.6	7.2	26.7	19.1	1.5
1–10 May	24.5	16.5	3.2	19.4	19.1	9.5	19.6	20.7	8.8	22.6	24.3	9.9
11–20 May	19.5	15.4	2.6	28.1	25.4	9.5	25.0	22.9	10.3	17.3	17.2	5.7
21–31 May	22.6	19.1	6.5	25.0	23.0	10.1	24.6	26.5	13.4	18.0	21.0	10.3
1–10 June	25.3	23.3	9.6	22.4	18.1	7.7	27.3	23.3	7.4	25.6	23.5	12.4
11–20 June	26.1	25.8	12.4	22.9	26.1	14.3	25.3	27.1	14.0	22.5	23.1	12.7
21–30 June	29.2	28.8	17.2	29.8	29.2	17.4	27.6	27.5	16.6	26.0	24.3	13.4
1–10 July	25.6	23.2	11.5	24.2	26.5	15.8	24.5	27.4	15.3	24.1	27.1	15.6
11–20 July	26.2	28.9	16.7	30.7	28.9	14.7	28.1	31.4	18.9	28.2	25.5	13.4
21–31 July	23.5	26.2	14.7	27.4	26.0	12.6	24.9	24.2	13.1	23.4	25.3	13.4
1–10 Aug.	22.3	25.0	13.1	20.3	26.1	13.4	20.5	24.3	12.3	24.4	27.2	14.8
11–20 Aug.	14.3	21.3	12.8	23.5	26.0	14.1	19.8	25.0	13.5	21.3	25.2	12.9
21–31 Aug.	19.0	24.3	13.0	22.3	27.2	15.7	21.4	25.4	10.9	17.8	26.4	14.4
1–10 Sept.	17.4	21.5	10.0	23.6	24.8	9.0	18.9	20.0	5.0	18.7	26.0	13.7
11–20 Sept.	16.2	23.3	9.3	18.5	28.9	12.5	17.5	19.6	5.4	16.9	21.9	8.6
21–30 Sept.	16.7	19.6	7.0	16.6	21.2	8.6	15.8	18.1	6.2	14.7	17.4	5.2

mass at harvest and grain yield. However, when appropriate, other important variables such as harvest index, leaf area index, and phenological stages will be included in the discussion. Since the three cultivars did not significantly deviate from each other, it was concluded that model failures to predict were not cultivar specific. Therefore, data will not be presented by cultivar but as an average response of the three cultivars.

**Initial Evaluation of Unmodified Model**

The original model did well in prediction of grain yield and total biomass at harvest in 1998 (Simulation 1; Table 5). Initial simulations resulted in correct simulation of average grain yield and 2% underprediction of total biomass at harvest, indicating that the model was well calibrated when both cultivar and site (SLPF) had been previously calibrated. The maximum leaf area index ( $LAI_{max}$ ) was underpredicted by 3% (data not shown). However, harvest index was overpredicted by 3% (data not shown). Since cultivars traits were optimized across sites and planting dates, model failures to address envi-

ronmental factors could be anticipated for one site/date vs. another. Underpredictions of  $LAI_{max}$  and total dry matter were more prevalent for the early planting date at Arlington and for both planting dates at the colder Hancock location, suggesting that cool temperature or soil type sensitivity could be a significant problem for accurate simulations with the present version of the CROPGRO-soybean model. Processes that are sensitive to temperature during early season include photosynthesis and vegetative development (Boote et al., 1998). This suggested the need to re-evaluate the temperature coefficients affecting these processes.

**Model Parameters Modification**

**Photosynthesis Changes**

Modifying the temperature functions for leaf photosynthesis did not improve predictions since it increased RMSE despite an increase in the mean simulated biomass and grain yield. Decreasing the base temperature for leaf photosynthesis from 8 to 6°C increased the bio-

**Table 4. Mean daily solar radiation (SR) and mean daily maximum ( $T_{max}$ ) and minimum ( $T_{min}$ ) temperatures between April and September at Hancock from 1997 to 2000.**

Period	1997			1998			1999			2000		
	SR	$T_{max}$	$T_{min}$									
	MJ m <sup>-2</sup>	°C		MJ m <sup>-2</sup>	°C		MJ m <sup>-2</sup>	°C		MJ m <sup>-2</sup>	°C	
1–10 Apr.	21.0	11.0	-1.9	17.0	12.0	1.0	10.7	1.3	-5.4	19.0	10.5	-1.3
11–20 Apr.	17.2	9.4	-1.7	18.1	15.7	3.2	14.2	5.6	-3.3	10.5	9.7	0.6
21–30 Apr.	23.3	16.0	1.5	25.0	18.3	2.2	24.3	15.2	2.3	27.5	19.7	2.2
1–10 May	25.1	16.5	2.5	19.6	19.6	9.3	20.6	18.7	6.8	24.7	24.7	11.3
11–20 May	19.9	14.6	2.0	28.7	28.4	14.8	18.7	17.9	7.7	18.4	16.3	5.4
21–31 May	24.2	18.5	6.9	27.0	22.7	9.7	23.1	20.2	6.5	19.9	21.0	9.7
1–10 June	28.3	23.2	11.1	20.8	17.2	6.4	22.7	26.4	12.0	19.6	23.7	12.6
11–20 June	26.1	24.6	11.8	22.6	25.4	14.4	24.1	26.8	14.5	21.7	22.2	12.5
21–30 June	30.3	28.8	17.5	29.4	29.0	17.1	24.7	26.8	14.5	25.4	24.0	13.5
1–10 July	26.1	22.5	10.8	25.2	26.2	15.1	26.4	27.4	16.1	24.4	26.1	16.2
11–20 July	27.6	29.2	17.5	30.4	29.5	15.6	23.6	28.9	19.5	26.1	25.4	13.6
21–31 July	22.9	26.0	14.8	27.9	26.6	13.9	23.3	28.9	19.5	23.0	25.5	14.2
1–10 Aug.	22.1	25.5	13.1	17.2	25.4	16.7	21.5	24.0	13.8	23.7	26.5	15.7
11–20 Aug.	14.9	21.0	12.2	30.4	29.5	15.6	20.7	24.3	13.4	21.4	24.7	13.8
21–31 Aug.	17.7	23.4	11.8	30.4	29.4	15.5	20.1	25.0	14.3	19.5	26.9	15.7
1–10 Sept.	18.5	21.9	11.0	21.9	25.4	9.7	19.5	25.1	12.7	17.7	23.4	13.2
11–20 Sept.	15.3	23.1	10.9	16.9	27.4	14.6	17.0	19.6	5.9	16.4	21.7	9.3
21–30 Sept.	16.4	19.5	8.4	13.7	19.5	8.0	15.7	19.1	5.3	15.3	17.1	6.6

**Table 5. Biomass and grain yield simulated in 1998 by the original CROPGRO-Soybean model (Simulation 1) and modifications affecting processes of photosynthesis (Simulations 2–4), N mobilization (Simulation 5), vegetative expansion (Simulations 6–7), and pod set (Simulation 9) as well as a combinations of changes (Simulations 10–13) compared with the experimental averages over three cultivars across five management systems at two planting dates ( $n = 30$ ).**

Simulation	Biomass	$a^{\dagger}$	$b^{\dagger}$	RMSE $^{\ddagger}$	$d^{\S}$	Grain yield	$a$	$b$	RMSE	$d$
	kg ha $^{-1}$			kg ha $^{-1}$		kg ha $^{-1}$			kg ha $^{-1}$	
(1) Original model	6997	3619	0.475	734	0.669	4135	3630	0.122	410	0.502
Photosynthesis changes										
(2) $T_b - P_g^{\ \}$	7229	3852	0.475	746	0.667	4272	3702	0.138	431	0.502
(3) $T_{min} - P_g^{\#}$	7289	3838	0.485	750	0.665	4299	3709	0.142	433	0.501
(4) (2) + (3)	7518	4056	0.487	832	0.634	4431	3776	0.158	494	0.488
N mobilization change										
(5) N mobilization $^{\dagger\dagger}$	7112	3816	0.464	723	0.673	4230	3685	0.132	419	0.505
Vegetative expansion										
(6) $T_b - V_S^{\ddagger\ddagger}$	6724	3337	0.477	739	0.667	4001	3543	0.111	408	0.458
(7) $T_b - S_LA^{\S\S}$	6964	3618	0.471	728	0.669	4120	3627	0.119	407	0.500
(8) (6) + (7)	6695	3355	0.470	751	0.659	3988	3530	0.110	411	0.453
Pod set										
(9) $T_b - \text{pod set}^{\ \ \}$	7000	3267	0.525	764	0.670	4229	2443	0.431	385	0.632
Example of combinations of changes										
(10) (5) + (7)	7118	3345	0.531	757	0.676	4322	2382	0.469	425	0.601
(11) (7) + (9)	6967	3255	0.522	758	0.670	4210	2450	0.425	377	0.641
(12) (5) + (7) + (9)	7086	3348	0.526	745	0.679	4303	2396	0.460	414	0.611
(13) (11) + sin function $^{\#\#}$	7021	3324	0.520	707	0.689	4228	2235	0.481	362	0.678
Observed data	7108					4142				

$^{\dagger}$   $a$  and  $b$  values of linear regression of predicted vs. observed data.

$^{\ddagger}$  RMSE, root mean square error.

$^{\S}$   $d$ , index of agreement.

$^{\|\}$  Base temperature for leaf photosynthesis decreased from 8 to 6°C.

$^{\#}$  Threshold values for the function calculating  $T_{min}$  effect on the subsequent day's light-saturated photosynthesis rate were moved from 0 and 19°C to -2 and 17°C.

$^{\dagger\dagger}$  Delaying of NVSMOB from 0.35 to 0.30.

$^{\ddagger\ddagger}$  Base temperature for main-stem node appearance increased from 7 to 9°C.

$^{\S\S}$  Reduced effect of low temperature on leaf expansion: Relative specific leaf area at 12°C changed from 0.25 to 0.35.

$^{\|\|\}$  Base temperature for pod addition decreased from 14 to 10°C.

$^{\#\#}$  Changed temperature function for reproductive development from planting to emergence and from emergence to flower from linear to sinus function with same cardinal temperatures.

mass at harvest and grain yield by 232 and 137 kg ha $^{-1}$ , respectively (Simulation 2; Table 5). Another temperature function was changed, which allowed minimum temperature to affect the next day's light-saturated photosynthesis rate ( $LF_{max}$ ). This is an asymptotic function, which ranges from zero at 0°C increasing to 1.0 (no limitation on  $LF_{max}$ ) when the minimum night temperature achieves 19°C. The threshold values of this function were changed to -2 and 17°C, respectively. This change in minimum temperature acted similarly to the base temperature change by increasing RMSE and increasing biomass and yield at harvest by 292 and 164 kg ha $^{-1}$ , respectively (Simulation 3; Table 5). The combined modifications (Simulation 4; Table 5) increased RMSE for biomass and grain yield by 13 and 20%, respectively. Combining the two temperature modifications decreased the  $d$  values but improved the slopes of predicted vs. observed biomass and grain yield slightly. These changes were therefore not included in the modified model.

### Nitrogen Mobilization

Sexton et al. (1998) identified the need to delay foliar N mobilization later into active seed growth under cooler temperature, thereby maintaining leaf N concentration and leaf photosynthesis late in the season. This change by itself delayed the onset of rapid N mobilization and increased biomass and grain yield by 115 and 95 kg ha $^{-1}$ , respectively (Simulation 5; Table 5). Overall, the degree of model predictability did not change. The

RMSE for biomass decreased by 1% and increased for grain yield by 2%. The  $d$  values and the slopes of predicted vs. observed biomass and grain yield improved slightly.

### Vegetative Expansion

Temperature functions affecting rate of vegetative node expression and leaf area expansion were also evaluated. Sinclair et al. (1991) showed that base temperature for the rate of node expression in soybean lies between 7 and 9°C. Increasing base temperature for rate of main-stem node appearance from 7 to 9°C decreased biomass and grain yield by 273 and 134 kg ha $^{-1}$ , respectively. Root mean square error increased for biomass from 734 to 739 kg ha $^{-1}$  and decreased for grain yield from 410 to 408 kg ha $^{-1}$  (Simulation 6; Table 5). CROPGRO-soybean also has a temperature effect on relative leaf area expansion, by which the relative specific leaf area (SLA) of new leaves increases from 0.25 at 12°C to 1.00 at 22°C. Based on previous observation (Pedersen, 2002), leaf area expansion was changed to be less sensitive to temperature (relative SLA from 0.25 at 12°C up to 0.35 at 12°C). This change decreased biomass and grain yield by 33 and 15 kg ha $^{-1}$ , respectively (Simulation 7; Table 5). Root mean square error decreased by 6 and 3 kg ha $^{-1}$  for biomass and grain yield, respectively. The combined modifications (Simulation 8; Table 5) increased RMSE for biomass and grain yield by 17 and 1 kg ha $^{-1}$ , respectively.

**Table 6. Biomass and grain yield simulated for 1997, 1999, and 2000 by the original CROPGRO-Soybean model and modified model based on the 1998 data coefficients for Hardin, CX232, and Spansoy 250.**

Simulation	Biomass	<i>a</i> †	<i>b</i> †	RMSE‡	<i>d</i> §	Grain yield	<i>a</i>	<i>b</i>	RMSE	<i>d</i>
	kg ha <sup>-1</sup>			kg ha <sup>-1</sup>		kg ha <sup>-1</sup>			kg ha <sup>-1</sup>	
1997 original	5792	6069	-0.040	1661	0.343	3088	3414	-0.085	1079	0.332
1997 modified	5870	5827	0.006	1574	0.359	3451	3627	-0.046	800	0.248
Observed data	6871					3813				
1999 original	6825	6093	0.102	1113	0.368	3982	4355	-0.093	787	0.181
1999 modified	6924	6078	0.118	1069	0.389	4208	4553	-0.086	797	0.261
Observed data	7161					4014				
2000 original	6607	5136	0.237	869	0.565	3914	3337	0.169	706	0.512
2000 modified	6654	5282	0.221	900	0.556	4061	3706	0.104	846	0.464
Observed data	6195					3423				

† *a* and *b* values of linear regression of predicted vs. observed data.

‡ RMSE, root mean square error.

§ *d*, index of agreement.

**Pod Set**

Based on previous observation by Hume and Jackson (1981), the temperature functions for pod addition rate were evaluated. Decreasing base temperature for pod addition from 14 to 10°C increased grain yield by 94 kg ha<sup>-1</sup> and in addition decreased RMSE for grain yield by 6% (Simulation 9; Table 5). However, this change only increased biomass by 3 kg ha<sup>-1</sup> but increased RMSE by 4%. However, slope and *d* value improved significantly to improve the model’s predictability.

**Final Model**

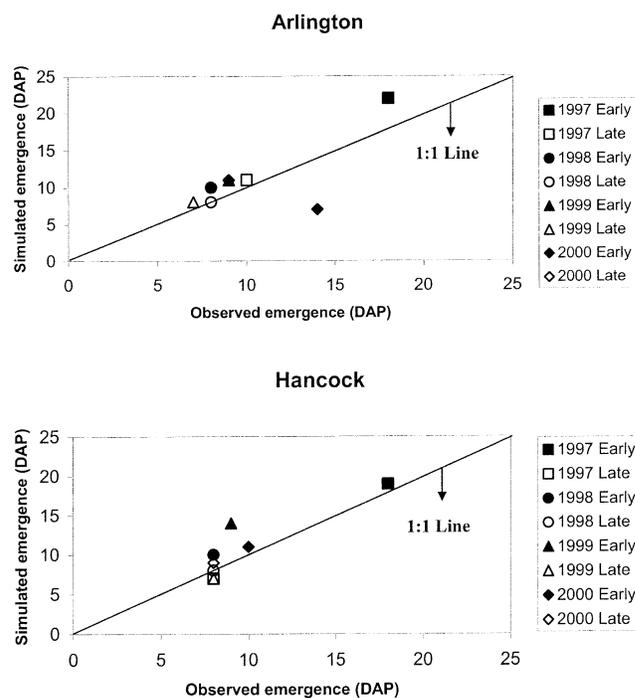
After having evaluated the individual processes, the temperature effect on leaf expansion and the base temperature for pod addition were combined with a change of the vegetative development rate curve from a linear function to a sinus function (Simulation 13; Table 5). This combination of modifications increased biomass and grain yield by 24 and 93 kg ha<sup>-1</sup>, respectively. The RMSE for biomass and grain yield decreased by 4 and 12%, respectively. In addition, *d* value and the slope of linear regression between predicted and observed data increased to improve the model’s predictability. These modified parameters improved the fit for the 1998 data and produced RMSE values in the same range as other observations from the central Corn Belt (Boote et al., 1997).

**Model Validation**

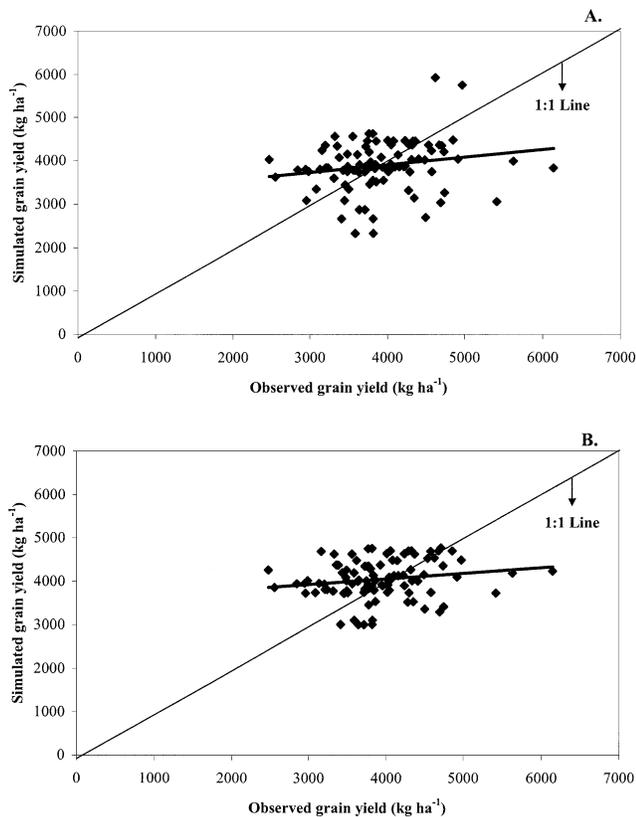
The original and the modified model were run with the independent 1997, 1999, and 2000 data to evaluate the degree of model improvement and predictability. The original model underpredicted biomass at harvest and grain yield by 16 and 19% and by 5 and 1% in 1997 and 1999, respectively (Table 6). However, in 2000, the original model overpredicted biomass at harvest and grain yield by 7 and 14%, respectively (Table 6). These are predictions using original model, with cultivar traits calibrated only to 1998. Root mean square error for biomass at harvest and grain yield was on average 1214 and 857 kg ha<sup>-1</sup>, respectively.

The modified parameter calibrated from the 1998 data set was then used for the validation of the model with the other 3 yr (1997, 1999, and 2000). Averaged across the 3 yr, modifications of the original model based

on the 1998 data increased biomass at harvest and grain yield slightly but with different degree of predictability. Since the model was fairly well calibrated for average total biomass and grain yield at harvest across cultivars, planting date, and locations in 1998, the variability observed with other weather years was speculated to be related with environmental effects among the 3 yr. In 1997, RMSE decreased by 5 and 26% for biomass at harvest and grain yield, respectively. In 1999, RMSE for biomass at harvest decreased by 4%, but RMSE for grain yield increased 1%. The picture was different for 2000 where the modified model resulted in increased RMSE for biomass at harvest and grain yield by 4 and 20%, respectively. An explanation for this is that May and June were exceptionally wet (254 mm of precipitation above the 20-yr average at Arlington), resulting in soil crusting, which delayed emergence and resulted in an underprediction by 7 d (Fig. 2). In addition, high inci-



**Fig. 2. Comparison of simulated vs. observed emergence for early and late planting dates at Arlington and Hancock (1997–2000). Data are averaged across four management systems and three cultivars at Arlington and across three cultivars at Hancock. DAP, days after planting.**

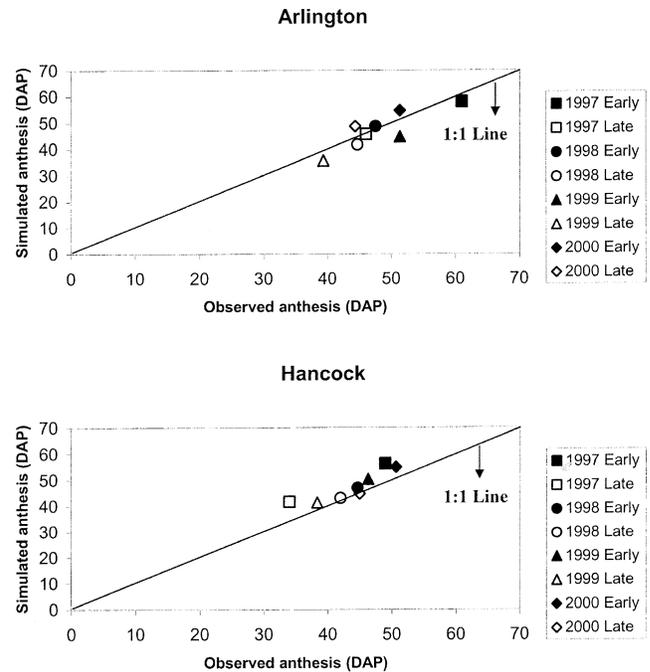


**Fig. 3.** Comparison of simulated vs. observed grain yield for (A) the original and (B) the modified model across management systems and planting dates in 1997–2000. There are no significant differences between the two regression slopes.

dence of *Sclerotinia* stem rot (caused by *Sclerotinia sclerotiorum*) was observed in 2001 (C.R. Grau, personal communication, 2003), which the model is not able to take into account.

The modified model overall did well predicting biomass at harvest and grain yield. Despite deviating by 19% in 2000 and an increase in RMSE in 1999 and 2000 (Table 6), grain yield differences between observed and simulated values improved from the original model, but the linear regression analysis did not produce a better fit as evidenced from no significant differences in the slope (Fig. 3). The model did not fully predict the range of variability in yield among the different management systems and planting dates because of the environmental effects between years. The model underestimation of the actual yield ranges was expected since the CROPGRO model does not allow or consider effects of tillage system, residue coverage, and disease incidence. CROPGRO overpredicted the lowest measured yields but underpredicted the highest (Fig. 3). A reason for this may be the high yields obtained for all years and the yield range in this study.

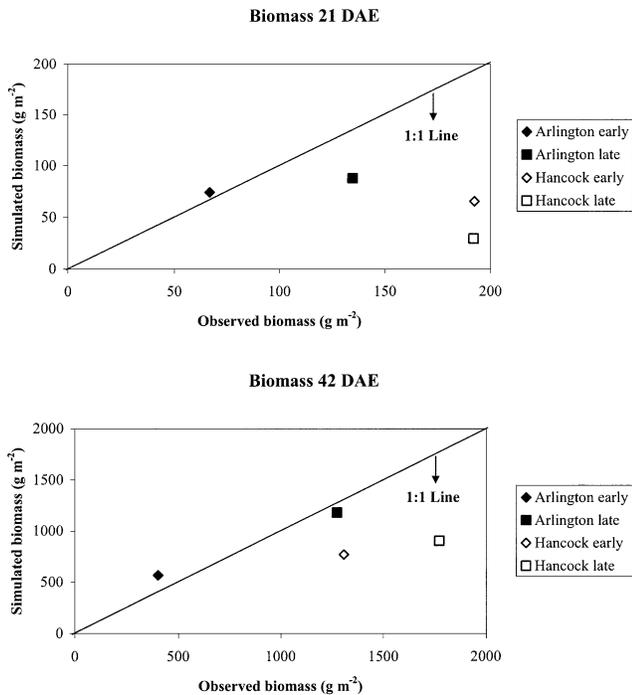
The time from planting to emergence and prediction of phenology may account for part of the high RMSE values in this data set (Sau et al., 1999; Sexton et al., 1998). Emergence was predicted well with a RMSE of 2.9 d across all planting dates and years. In all but three cases, simulated emergence was within 3 d of observed



**Fig. 4.** Comparison of simulated vs. observed anthesis for early and late planting dates at Arlington and Hancock (1997–2000). Data are averaged across four management systems and three cultivars at Arlington and across three cultivars at Hancock. DAP, days after planting.

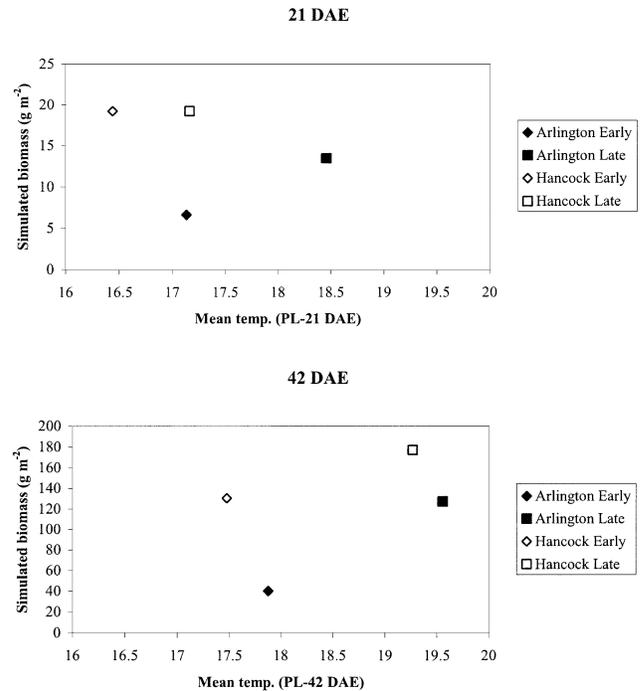
values (Fig. 2). The explanation for similarity among these three cases is that they all are from the early planting date. Emergence at both Arlington and Hancock was, in general, slightly overpredicted (late). Arlington 2000 was an exception compared with the other years because of a reduction in solar radiation (Table 3) and significant rainfall amounts during the emergence period. Most likely, the significant rainfall saturated the soil profile, cooling the soil and increasing the heat capacity of the upper layers. Thus, it would have required much more heat to warm the upper soil layers. Thus, the real crop was delayed, but the modeled crop does not cool the soil with excess rain or evaporation and was not delayed enough. Possible other sources of errors in predicting emergence could be inaccurate initial conditions at the start of simulations and wrong estimates of soil physical properties. Anthesis dates were predicted within 4 d with a RMSE of 2.5 d across planting dates and years. Anthesis dates at Arlington were predicted very well but were slightly overpredicted at Hancock (Fig. 4). The average differences between actual and simulated dates were 0.8 and 1.3 d for flowering at Arlington and Hancock, respectively.

The high RMSE values and the variability of observed slopes and intercepts indicate (Tables 5 and 6) that the model may not be able to simulate the actual site-specific and year-specific environmental variations. Given that soybean emergence is affected by soil temperature (Andales et al., 2000) and temperature functions in the CROPGRO model after emergence are responsive mainly to air temperature (Hoogenboom et al., 1992), variability and inaccurate predictability for these data could also be a result of a lack of sensitivity to soil temperature



**Fig. 5.** Comparison of simulated vs. observed biomass yield for early and late planting dates at Arlington and Hancock in 1998. Data are averaged across four management systems and three cultivars at Arlington and across three cultivars at Hancock. DAE, days after emergence.

for different soil types and planting dates. Total above-ground biomass during early vegetative growth was evaluated for the first two sampling dates (21 and 42 DAE) for the 1998 data set (Fig. 5). At both sampling dates, biomass yield was underpredicted for both planting dates at Hancock and for the late planting date at Arlington whereas biomass for the early planting date at Arlington was slightly overpredicted. Similar trends were observed for leaf area index (data not shown). An improvement in predictability was observed at 42 DAE compared with 21 DAE, suggesting that temperature in one way or another may account for this underprediction. It would appear that the cooler air temperatures for the Hancock site vs. Arlington site dominated the model simulations; however, the actual soil was likely warmer for the Hancock site because of the sandy loam soil. We conclude the model needs modifications to consider effects of soil temperature on early growth and effects of soil texture, residue cover, and soil evaporation (wetness) on soil temperature. Even though yield variation for early and late planting date was reproduced well by the crop model for each location, the model inaccurately predicted early vegetative growth because of the effect of site-specific and planting date-specific differences on biomass accumulation (Fig. 6). Highest biomass yield was observed at both sampling dates at Hancock despite a lower mean-averaged air temperature. An increase in mean air temperature had a greater effect on biomass yield at Arlington than at Hancock. An explanation for a higher biomass accumulation at Hancock could be that the lower water-holding capacity in a sandy loam soil reduced the estimated heat capacity



**Fig. 6.** Simulated biomass yield at 21 and 42 d after emergence (DAE) plotted against mean air temperature from planting (PL) to emergence for early and late planting dates at Arlington and Hancock in 1998. Data are averaged across four management systems and three cultivars at Arlington and across three cultivars at Hancock.

of the soil and therefore required less heat input to maintain the temperature of the upper soil layers, which the model does not consider.

## CONCLUSION

Analyzing cultivar performance under various management systems and environmental conditions can serve as help for evaluating model performance for a specific region. The CROPGRO-Soybean model predicted emergence and anthesis within a 3- to 8-d range, respectively. The original model did well in prediction of grain yield and total biomass at harvest, only underpredicting by 111 and 7 kg ha<sup>-1</sup>, respectively.

Modifications of temperature functions influencing leaf expansion and base temperature for pod addition improved the CROPGRO-soybean prediction error in 1998. However, these changes were not found to be beneficial for the three independent years used for model validation, and thus we reject our hypothesis that refinements in temperature functions in CROPGRO-Soybean could be made to facilitate model use across environments in Wisconsin. The inaccuracy of the model across years and locations is attributed to somewhat late prediction of emergence on average but too early emergence for problem fields where excess rain occurred. This suggests a need for greater temperature sensitivity to soil types. There was general underprediction of leaf area index and biomass accumulation (emergence to 42 DAE) for sites that had sandy soils that warmed quickly. This supports our hypothesis that the model underpredicts growth and yield during vegetative growth at cooler air temperature environments.

This study was an effort to improve the CROPGRO-Soybean model for cooler environments in the upper Midwest. It is suggested that some temperature functions may need to be linked with improved soil temperature predictions rather than air temperature to improve prediction of emergence and early growth of soybean for cooler environments. Linkage of soil temperature effects on early vegetative growth and leaf expansion rate may warrant further consideration in the future improvement of the CROPGRO-model once soil temperature is correctly predicted and considers soil texture and rainfall-irradiance environment.

### ACKNOWLEDGMENTS

The authors thank Dr. Edward S. Oplinger for assistance early in the development of this study and John M. Gaska, Mark Martinka, and Kathy Bures for their technical assistance. This research was partially funded by Iowa Soybean Promotion Board, Illinois Soybean Checkoff Board, and Wisconsin Soybean Marketing Board.

### REFERENCES

- Allen, E.M., W.D. Batchelor, and T.S. Colvin. 1996. Validation of corn and soybean models in Iowa: Implications for precision farming. ASAE paper 96-1006. ASAE, St. Joseph, MI.
- Andales, A.A., W.D. Batchelor, and C.E. Anderson. 2000. Modification of a soybean model to improve soil temperature and emergence date prediction. *Trans. ASAE* 43:121-129.
- Boote, K.J., J.W. Jones, G. Hoogenboom, and N.B. Pickering. 1998. The CROPGRO model for grain legumes. p. 99-128. *In* G.Y. Tsuji et al. (ed.) *Understanding options for agricultural production*. Kluwer Academic Publ., Boston.
- Boote, K.J., J.W. Jones, G. Hoogenboom, and G.G. Wilkerson. 1997. Evaluation of the CROPGRO-Soybean model over a wide range of experiments. p. 113-133. *In* Kropff et al. (ed.) *Systems approaches for sustainable agricultural development: Applications of systems approaches at the field level*. Kluwer Academic Publ., Boston.
- Boote, K.J., M.J. Kropff, and P.S. Bindraban. 2001. Physiology and modeling of traits in crop plants: Implications for genetic improvement. *Agric. Syst.* 70:395-420.
- Boote, K.J., and M. Tollenaar. 1994. Modeling genetic yield potential. p. 533-565. *In* K.J. Boote et al. (ed.) *Physiology and determination of crop yield*. ASA, CSSA, and SSSA, Madison, WI.
- Caufield, F., and J.A. Bunce. 1988. Comparative responses of photosynthesis to growth temperature in soybean (*Glycine max* [L.] Merrill) cultivars. *Can. J. Plant Sci.* 68:419-425.
- Cooper, R.L. 1977. Response of soybean cultivars to narrow rows and planting rates under weed-free conditions. *Agron. J.* 69:89-92.
- Curry, R.B., J.W. Jones, K.J. Boote, R.M. Peart, L.H. Allen, Jr., and N.B. Pickering. 1995. Responses of soybean to predicted climate change in the USA. p. 163-182. *In* C. Rosenzweig et al. (ed.) *Climate change and agriculture: Analysis of potential international impacts*. ASA Spec. Publ. 59. ASA, Madison, WI.
- Egli, D.B., and W. Bruening. 1992. Planting date and soybean yield: Evaluation of environmental effects with a crop simulation model SOYGRO. *Agric. For. Meteorol.* 62:19-29.
- Fehr, W.R., and C.E. Caviness. 1977. Stages of soybean development. Spec. Rep. 80. Iowa Agric. Home Econ. Exp. Stn., Iowa State Univ., Ames.
- Grant, R.F. 1989. Simulation of carbon assimilation and partitioning in maize. *Agron. J.* 81:563-571.
- Grimm, S.S., J.W. Jones, K.J. Boote, and D.C. Herzog. 1994. Modeling the occurrence of reproductive stages after flowering for four soybean cultivars. *Agron. J.* 86:31-38.
- Harley, P.C., J.A. Weber, and D.M. Gates. 1985. Interactive effects of light, leaf temperature, CO<sub>2</sub> and O<sub>2</sub> on photosynthesis in soybean. *Planta* 165:249-263.
- Hoogenboom, G., J.W. Jones, and K.J. Boote. 1992. Modeling growth, development, and legumes using SOYGRO, PNUTGRO, and BEANGRO: A review. *Trans. ASAE* 35:2043-2056.
- Hoogenboom, G., J.W. White, J.W. Jones, and K.J. Boote. 1994. BEANGRO, a process-oriented dry bean model with a versatile user interface. *Agron. J.* 86:182-190.
- Hume, D.J., and K.H. Jackson. 1981. Pod formation in soybeans at low temperature. *Crop Sci.* 21:933-937.
- Hunt, L.A., S. Pararajasingham, J.W. Jones, G. Hoogenboom, D.T. Imamura, and R.M. Ogoshi. 1993. GENCALC—software to facilitate the use of crop models for analyzing field experiments. *Agron. J.* 87:1090-1094.
- Jones, J.W. 1993. Decision support systems for agricultural development. p. 459-471. *In* Penning de Vries et al. (ed.) *Systems approaches for agricultural development*. Kluwer Academic Publ., Boston.
- Jones, J.W., L.A. Hunt, G. Hoogenboom, D.C. Godwin, U. Singh, G.Y. Tsuji, N.B. Pickering, P.K. Thornton, W.T. Bowen, K.J. Boote, and J.T. Ritchie. 1994. Input and output files. p. 1-94. *In* Tsuji et al. (ed.) *Decision Support System for Agrotechnology Transfer (DSSAT) Version 3. Vol. 2*. Univ. of Hawaii, Honolulu.
- Jones, P.N., and P.S. Carberry. 1994. A technique to develop and validate simulation models. *Agric. Syst.* 46:427-442.
- Marowitch, J., G.R. Pautsch, and J. Hoddinott. 1986. The influence of plant temperature on photosynthesis and translocation rates in bean and soybean. *Can. J. Bot.* 64:2337-2342.
- Mavromatis, T., K.J. Boote, J.W. Jones, A. Irmak, D. Shinde, and G. Hoogenboom. 2001. Developing genetic coefficients for crop simulation models with data from crop performance trials. *Crop Sci.* 41:40-51.
- O'Leary, G.J., D.J. Connor, and D.H. White. 1985. A simulation model of the development, growth and yield of the wheat crop. *Agric. Syst.* 17:1-26.
- Paz, J.O., W.D. Batchelor, T.S. Colvin, S.D. Logsdon, T.C. Kaspar, and D.L. Karlen. 1998. Calibration of a crop growth model to predict spatial yield variability. *Trans. ASAE* 41:1527-1534.
- Pedersen, P. 2002. Yield potential of soybean under various management systems. Ph.D. diss. Univ. of Wisconsin, Madison.
- Pedersen, P., and J.G. Lauer. 2003. Soybean agronomic response to management systems in the upper Midwest. *Agron. J.* 95:1146-1151.
- Piper, E.L., K.J. Boote, J.W. Jones, and S.S. Grimm. 1996. Comparison of two phenology models for predicting flowering and maturity date of soybean. *Crop Sci.* 36:1606-1614.
- Sau, F., K.J. Boote, and B. Ruiz-Nogueira. 1999. Evaluation and improvement of CROPGRO-soybean model for a cool environment in Galicia, northwest Spain. *Field Crops Res.* 61:273-291.
- Schulte, E.E., and L.M. Walsh. 1994. Soil water. p. 21-28. *In* L.G. Deith (ed.) *Management of Wisconsin soils*. A3588. Univ. of Wisconsin Coop. Ext. Publ., Madison.
- Seddigh, M., G.D. Jolliff, and J.H. Orf. 1989. Night temperature effects on soybean phenology. *Crop Sci.* 29:400-406.
- Sexton, P.J., W.D. Batchelor, K.J. Boote, and R. Shibles. 1998. Evaluation of CROPGRO for prediction of soybean nitrogen balance in a midwestern environment. *Trans. ASAE* 41:1543-1548.
- Sexton, P.J., J.W. White, and K.J. Boote. 1994. Yield-determining processes in relation to cultivar seed-size of common bean. *Crop Sci.* 34:84-91.
- Sinclair, T.R., S. Kitani, K. Hinson, J. Bruniard, and T. Horie. 1991. Soybean flowering date: Linear and logistic models based on temperature and photoperiod. *Crop Sci.* 31:786-790.
- Wallach, D., and B. Goffinet. 1987. Mean square error of prediction in models for studying ecological and agronomic systems. *Biometrics* 43:561-573.
- Wilkerson, G.G., J.W. Jones, K.J. Boote, K.T. Ingram, and J.W. Mishoe. 1983. Modeling soybean growth for crop management. *Trans. ASAE* 26:63-73.
- Willmott, C.J. 1982. Some comments on the evaluation of model performance. *Bull. Am. Meteorol. Soc.* 63:1309-1313.
- Willmott, C.J., and D.E. Wicks. 1980. An empirical method for the spatial interpolation of monthly precipitation within California. *Phys. Geogr.* 1:59-73.