in silico
Plants
OXFORD

Soybean-BioCro: a semi-mechanistic model of soybean growth

Megan L. Matthews[1,2,*], Amy Marshall-Colón[2,3,4], Justin M. McGrath[2,3,5],
Edward B. Lochocki[2] and Stephen P. Long[2,3,4,6]

[1]Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
[2]Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
[3]Department of Plant Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
[4]Institute for Sustainability, Energy, and Environment, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
[5]USDA ARS Global Change and Photosynthesis Research Unit, Urbana, IL 61801, USA
[6]Department of Crop Sciences, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
*Corresponding author's e-mail address: mlmatth2@illinois.edu

Handling Editor: Graeme Hammer

ABSTRACT

Soybean is a major global source of protein and oil. Understanding how soybean crops will respond to the changing climate and identifying the responsible molecular machinery are important for facilitating bioengineering and breeding to meet the growing global food demand. The BioCro family of crop models are semi-mechanistic models scaling from biochemistry to whole crop growth and yield. BioCro was previously parameterized and proved effective for the biomass crops Miscanthus, coppice willow and Brazilian sugarcane. Here, we present Soybean-BioCro, the first food crop to be parameterized for BioCro. Two new module sets were incorporated into the BioCro framework describing the rate of soybean development and carbon partitioning and senescence. The model was parameterized using field measurements collected over the 2002 and 2005 growing seasons at the open air $[CO_2]$ enrichment (SoyFACE) facility under ambient atmospheric $[CO_2]$. We demonstrate that Soybean-BioCro successfully predicted how elevated $[CO_2]$ impacted field-grown soybean growth without a need for re-parameterization, by predicting soybean growth under elevated atmospheric $[CO_2]$ during the 2002 and 2005 growing seasons, and under both ambient and elevated $[CO_2]$ for the 2004 and 2006 growing seasons. Soybean-BioCro provides a useful foundational framework for incorporating additional primary and secondary metabolic processes or gene regulatory mechanisms that can further aid our understanding of how future soybean growth will be impacted by climate change.

KEYWORDS: Climate change; crop modelling; elevated $[CO_2]$; semi-mechanistic modelling, soybean.

1. INTRODUCTION

Soybean (*Glycine max*) is the fourth most important seed crop in terms of global production with major production areas in North America, South America and Asia (Pagano and Miransari 2016). Soybeans have the highest protein content and second highest oil content among the major food crops (Nafziger 2016). In addition to being a popular protein source in human diets, soybean is a major component of livestock and aquaculture feed, vegetable oil and biodiesel (United Soybean Board 2020). Future soybean production is dependent on the ability of soybean crops to adapt to climate change and the associated abiotic

and biotic stressors. Plants are complex, sessile organisms that survive using internal regulatory mechanisms and machinery to respond to stressors and resources that impact plant growth and health such as rising $[CO_2]$ levels, drought, flooding, decreased and elevated temperatures, nutrient deficiency and nutrient sufficiency (Musser *et al.* 1983; Seddigh *et al.* 1989; Srinivasan and Arihara 1994; Kurosaki and Yumoto 2003; Das *et al.* 2016; Haak *et al.* 2017; Coutinho *et al.* 2018; Jähne *et al.* 2019).

As a limiting substrate for photosynthesis, rising atmospheric $[CO_2]$ is expected to have a large impact on crop development and

yield (Ainsworth *et al.* 2007; Wang *et al.* 2008; Bishop *et al.* 2015). The rate of increase of atmospheric $[CO_2]$ year to year was approximately 0.9 ppm per year in the 1960s rising to an average of 2.4 ppm per year from 2009 to 2019, resulting in 415 ppm in 2020 (Tans and Keeling 2021). Atmospheric $[CO_2]$ is expected to continue rising with projections ranging from 800 to more than 1100 ppm by the year 2100 (Collins *et al.* 2013). Global temperatures are also rising and becoming more variable leading to more frequent extreme heat and other weather events (Herring *et al.* 2018).

Understanding how soybean crops will respond to these abiotic stressors and identifying the molecular machinery causing these responses are important tasks for sustainable growth of soybean with increased demand and improved nutritional content. Studies exploring soybean responses and behaviour have been carried out at several biological levels of organization: genetic (Hyten *et al.* 2010; Lam *et al.* 2010; Qiu *et al.* 2013; Song *et al.* 2015; Zhou *et al.* 2015), transcriptomic (Choi *et al.* 2007; Leakey *et al.* 2009), proteomic (Das *et al.* 2016; Muñoz *et al.* 2016), metabolomic (Das *et al.* 2017; Coutinho *et al.* 2018) and phenotypic (Morgan *et al.* 2005). Pooling these resources may help to further explore how soybean crops will respond to a changing climate and identify potential strategies to improve soybean growth, yield and nutrition under these scenarios.

Computational models are one tool used to explore how plants will respond under different stress conditions. Once developed, computational models can be used to test a number of different scenarios that would otherwise be time or cost-prohibitive to perform experimentally (Prusinkiewicz and Runions 2012). Computational models can also help uncover new insights about the underlying machinery of biological systems (Prusinkiewicz and Runions 2012). There are multiple models to predict the growth of our major crops, including soybean. Current models, however, are generally based on the wealth of empirical data collected across years and geographical locations. These empirical relationships, however, are not designed for future climate scenarios where crop response to higher atmospheric $[CO_2]$, interacting with changes in temperature, precipitation, and other soil and climate factors, is not known. Understanding how crops will respond under these climate scenarios requires predicting outside of experience, which is best done by capturing the known mechanisms by which plants respond to rising $[CO_2]$, temperature and other variables in combination. BioCro is designed to accommodate such mechanisms. Mechanistic models of crop processes improve our ability to explore how a changing environment may impact crop growth, and can help identify strategies for engineering crops for improved performance under these scenarios (Kannan *et al.* 2019). Mechanistic models often require large amounts of information that can be difficult to obtain. However, by creating a modular framework, we can integrate mechanistic descriptions of crop model dynamics, created by domain experts, as they become developed (Marshall-Colon *et al.* 2017). Such domains include photosynthesis, respiration, partitioning, phenology and whole plant metabolism among others.

BioCro (Miguez *et al.* 2009) is a modular, semi-mechanistic dynamic crop growth model framework simulated using site-specific soil properties and hourly weather measurements including solar radiation, wind speed, precipitation, temperature and humidity. BioCro is based on WIMOVAC (Humphries and Long 1995; Humphries 2002)

and built on underlying biophysical and biochemical photosynthesis mechanisms that respond to changes in atmospheric $[CO_2]$, temperature (Bernacchi *et al.* 2001, 2003; Yang *et al.* 2016) and water availability. Further, BioCro is designed for easy expansion to incorporate new and more informed modules.

Previously BioCro has been parameterized for the C_4 biomass crops *Miscanthus × giganteus* (Miguez *et al.* 2009), *Panicum virgatum* (Miguez *et al.* 2012), Brazilian sugarcane (Jaiswal *et al.* 2017) and C_3 coppiced willow (Wang *et al.* 2015). BioCro has been used to spatially explore crop growth over large areas such as *M. × giganteus* and *P. virgatum* growth in the conterminous USA (Miguez *et al.* 2012), *M. × giganteus* and coppice willow growth in Denmark (Larsen *et al.* 2016) and Brazilian sugarcane in Brazil (Jaiswal *et al.* 2017). Larsen *et al.* found that soil characteristics were an important factor to whether Miscanthus or willow would be more productive in different regions of Denmark, and that variation in soil moisture was a more important contributor to yield than radiation or precipitation (Larsen *et al.* 2016). Jaiswal *et al.* applied BioCro to project sugarcane growth across Brazil using five major global circulation climate change models and multiple land use scenarios, and evaluated sugarcane's applicability as an alternative to crude oil (Jaiswal *et al.* 2017).

Here, we present Soybean-BioCro, the first food crop to be modeled using the BioCro family of crop growth models. We incorporated two new sets of modules into BioCro: (i) carbon partitioning and senescence modules using logistic functions based on the framework used in JULES-crop (Osborne *et al.* 2015); and (ii) a soybean development rate module based on selected functions from the SOYDEV model (Setiyono *et al.* 2007). The logistic equation-based partitioning module works on the same principles as a carbon partitioning table commonly used in crop growth models. These partitioning tables define a set percentage of the net carbon assimilated to be allocated to each of the crop organs at a given time. Typically, partitioning tables have used piecewise constant functions to define the fraction of carbon allocated to each crop organ throughout a growing season. To adequately capture the non-linear carbon allocation strategies across a growing season, however, can require many piecewise components, increasing the number of parameters needed. Logistic functions can capture this behaviour using fewer parameters. We added a specific soybean development rate module for Soybean-BioCro since soybean growth is dependent on both temperature and night length (Setiyono *et al.* 2007). As such, the accumulated thermal time that has been used as a proxy for crop development in other BioCro crops, which were non-flowering or allocated only a tiny fraction of biomass to seed, was not adequate for predicting soybean development across different growing seasons.

These modules resulted in improved parameterization for carbon partitioning and biomass senescence, and improved development stage prediction needed for accurate partitioning and senescence across different years. Partitioning and senescence parameters were determined using an evolutionary optimization algorithm to fit the predicted leaf, stem and pod biomass results to experimentally measured biomasses for the cultivar Pioneer 93B15 under ambient atmospheric $[CO_2]$ over the 2002 and 2005 growing seasons at the SoyFACE facility in Urbana, IL (Bernacchi *et al.* 2005; Morgan *et al.* 2005; Bishop *et al.* 2015). From this we successfully predicted the leaf, stem and pod biomasses

in ambient $[CO_2]$ (2004, 2006) and elevated $[CO_2]$ (2002, 2004–06), demonstrating that Soybean-BioCro accurately predicts leaf, stem and pod biomasses under elevated $[CO_2]$ conditions from mechanistic photosynthetic principles.

2. METHODS
2.1 BioCro

The core modules of the BioCro family of crop models include canopy photosynthesis, canopy energy balance, soil-water processes, crop development rate, carbon allocation and senescence. Soybean-BioCro uses several previously developed BioCro modules (Humphries 2002; Miguez *et al.* 2009, 2012; Wang *et al.* 2015; Larsen *et al.* 2016; Jaiswal *et al.* 2017), including (i) a 10-layer sun-shade canopy model that incorporates the light interception, humidity and wind speed profiles throughout the canopy layers; (ii) C_3 photosynthesis and canopy energy balance models and (iii) a two-layer soil profile model that uses site-specific soil properties to calculate the water available for growth and the impact of water stress on stomatal function. In addition to these modules, we added two new module sets for Soybean-BioCro, a soybean development rate module based on the SOYDEV model (Setiyono *et al.* 2007), and logistic equation partitioning growth and senescence modules based on the partitioning growth framework used in the JULES-crop model (Osborne *et al.* 2015). The two new BioCro module sets are described in the following sections. The equations, parameters and initial values used in all of the Soybean-BioCro modules are described in **Supporting Information—Text S1**.

2.1.1 Soybean development rate module.
The soybean development rate module is split into five phases that calculate development rate from (i) sowing to emergence (VE), (ii) emergence to the cotyledon stage (V0), (iii) the cotyledon stage to the end of floral induction (R0), (iv) the end of floral induction to flowering (R1) and (v) flowering to maturity (R7). We mapped these phases to a crop neutral development index (DVI) used in the partitioning growth module, with

$$
\begin{aligned}
-1 \leq \text{DVI} < 0 : & \quad \text{Sowing} - \text{VE} \\
0 \leq \text{DVI} < \tfrac{1}{3} : & \quad \text{VE} - \text{V0} \\
\tfrac{1}{3} \leq \text{DVI} < \tfrac{2}{3} : & \quad \text{V0} - \text{R0} \\
\tfrac{2}{3} \leq \text{DVI} < 1 : & \quad \text{R0} - \text{R1} \\
1 \leq \text{DVI} : & \quad \text{R1} - \text{R7}.
\end{aligned}
$$

2.1.1.1 Sowing to emergence: $-1 \leq DVI < 0$.
We used accumulated thermal time to calculate the development rate for phase (i) sowing to emergence.

$$
r = \frac{temp - T_b}{TT_{\text{emr}}} \tag{1}
$$

where r is the soybean development rate, *temp* is the air temperature which we use as a proxy for soil temperature, T_b is the base or minimum air temperature necessary for soybean to grow and TT_{emr} is the accumulated thermal time at which emergence occurs.

2.1.1.2 Emergence to maturity: $0 \leq DVI.$
From phases (ii) emergence to (v) maturity we used a subset of the SOYDEV photothermal functions (Equations (2)–(7)) (Setiyono *et al.* 2007).

$$
r = R_{\max} f_T f_P \tag{2}
$$

where,

$$
f_T(t, T_{min}, T_{opt}, T_{max}) = \begin{cases} \frac{2(t-T_{min})^\alpha (T_{opt}-T_{min})^\alpha - (t-T_{min})^{2\alpha}}{(T_{opt}-T_{min})^{2\alpha}}, & \text{if } T_{min} < t < T_{max}. \\ 0, & \text{else.} \end{cases} \tag{3}
$$

$$
\alpha = \frac{ln(2)}{ln\left(\frac{T_{max}-T_{min}}{T_{opt}-T_{min}}\right)} \tag{4}
$$

$$
f_P(p, P_{opt}, P_{crit}) = \begin{cases} \left[\left(\left(1 + \frac{p-P_{opt}}{m}\right)\left(\frac{P_{crit}-p}{P_{crit}-P_{opt}}\right)\right)^\gamma\right]^\beta, & \text{if } P_{opt} \leq p \leq P_{crit}. \\ 1, & \text{if } p < P_{opt}. \\ 0, & \text{if } p > P_{crit}. \end{cases} \tag{5}
$$

$$
\beta = \frac{ln(2)}{ln\left(1 + \frac{P_{crit}-P_{opt}}{m}\right)} \tag{6}
$$

$$
\gamma = \frac{P_{crit} - P_{opt}}{m} \tag{7}
$$

$R_{\max}$ is the maximum development rate. The temperature response function f_T scales the development rate as a function of temperature. At the optimal temperature, T_{opt}, $f_T = 1$. When the temperature is below T_{min} or above T_{max} then $f_T = 0$ and therefore the development rate $r = 0$. The photoperiod response function f_P scales the development rate as a function of day length. When the day length is greater than the critical day length, P_{crit}, $f_P = 0$ and the development rate $r = 0$. When the day length is less than or equal to the optimal day length for development, P_{opt}, then $f_P = 1$. For some phases, the development rate is dependent on only temperature ($r = R_{\max} f_T$) or only day length ($r = R_{\max} f_P$) as defined in (Setiyono *et al.* 2007). In our simulations, we used the SOYDEV temperature and photoperiod parameters for maturity group 3 soybean cultivars (Setiyono *et al.*, 2007). A complete list of all of the parameters and equations for the five developmental phases are included in **Supporting Information—Text S1** (Table S6, Eqs S82–S101).

2.1.2 Logistic equation-based carbon partitioning and senescence modules.
We created a new set of carbon partitioning modules based on the logistic-based functions from the JULES-crop land simulation model (Osborne *et al.*, 2015).

$$
k_{\text{Root}} = \frac{exp(\alpha_R + \beta_R x)}{exp(\alpha_R + \beta_R x) + exp(\alpha_L + \beta_L x) + exp(\alpha_S + \beta_S x) + 1} \tag{8}
$$

$$k_{\text{Leaf}} = \frac{exp(\alpha_L + \beta_L x)}{exp(\alpha_R + \beta_R x) + exp(\alpha_L + \beta_L x) + exp(\alpha_S + \beta_S x) + 1} \tag{9}$$

$$k_{\text{Stem}} = \frac{exp(\alpha_S + \beta_S x)}{exp(\alpha_R + \beta_R x) + exp(\alpha_L + \beta_L x) + exp(\alpha_S + \beta_S x) + 1} \tag{10}$$

$$k_{\text{Pod}} = \frac{1}{exp(\alpha_R + \beta_R x) + exp(\alpha_L + \beta_L x) + exp(\alpha_S + \beta_S x) + 1} \tag{11}$$

where k_{Root}, k_{Leaf}, k_{Stem} and k_{Pod} represent the fraction of the net carbon assimilated that is allocated to each of the four plant organs over the growing period. Osborne *et al.* used a developmental index (DVI) that ranged from -1 to 2, representing emergence, vegetative and reproductive development periods, as the dependent variable, x, in their implementation of this partitioning growth model (Osborne *et al.* 2015). Other metrics such as accumulated thermal time or crop-specific development stages can also be used. For Soybean-BioCro, we used a DVI range from -1 to $2+$ that mapped to specific development phases as predicted by the soybean development rate module (Section 2.1.1). The α_R, β_R, α_L, β_L, α_S, β_S parameters define the shape of the partitioning functions over the growing period (Fig. 1C).

We used a similar approach to model leaf and stem senescence. We determined what fraction of the leaf, s_{Leaf}, and stem, s_{Stem}, biomasses that are senesced at each point in the growing season using the

switch-like logistic function (Equations (12) and (13)). r_{sL} is the maximum fraction of leaf biomass that will be senesced at a given point. and α_{sL} and β_{sL} determine when the leaf starts to senesce. r_{sS}, α_{sS} and β_{sS} are the corresponding parameters for stem senescence. In our simulations, we assumed that the root and pod do not senesce.

$$s_{\text{Leaf}} = \frac{r_{sL}}{1 + exp(\alpha_{sL} + \beta_{sL} x)} \tag{12}$$

$$s_{\text{Stem}} = \frac{r_{sS}}{1 + exp(\alpha_{sS} + \beta_{sS} x)} \tag{13}$$

2.2 Parameter optimization

The parameters for carbon partitioning (α_R, β_R, α_L, β_L, α_S, β_S, Equations (8)–(11)) and senescence (r_{sL}, α_{sL}, β_{sL}, r_{sS}, α_{sS}, β_{sS}, Equations (12) and (13)) were obtained using the DEOptim function (Mullen *et al.* 2011) to minimize the weighted mean square errors (MSEs) between the predicted and measured biomasses for the 2002 and 2005 growing seasons at SoyFACE under ambient atmospheric $[CO_2]$ (Equations (14)–(16)).

For each shoot tissue (leaf, stem and pod), we squared the difference between the predicted biomass $\hat{y}$ and the average of the measured biomasses for that tissue, $\bar{y}$, for each of the N days where the biomass was experimentally measured. For the 2002 and 2005 growing seasons, the leaf, stem and pod biomasses were measured $N = 8$ times throughout the growing seasons. We scaled this error term by the maximum of
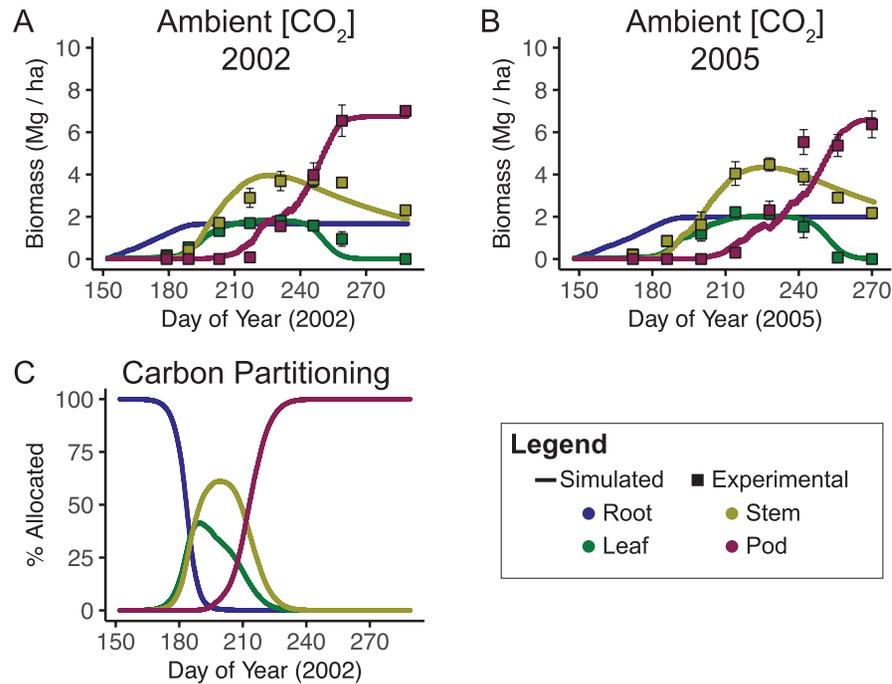

Figure 1. Simulated root, leaf, stem and pod biomasses and the experimental leaf, stem and pod biomasses used for parameter fitting, under ambient atmospheric $[CO_2]$ in (A) 2002 and (B) 2005. (C) Fraction of carbon allocated to the four crop organs over the 2002 growing season.

the measured biomass, $max(\bar{\mathbf{y}})$, for a given tissue to remove bias in the optimization algorithm due to the different ranges of the leaf, stem and pod biomasses. We took the average of these squared errors to calculate the MSE for the leaf, stem and pod tissues (Equation (15)).

As root biomass was not measured, we assumed the root biomass would grow to 17 % of the maximum of the total measured shoot biomass (Ordóñez *et al.* 2020). The error term for root biomass (Equation (16)) is the squared error between the maximum predicted root biomass, $max(\hat{\mathbf{y}}_{Root})$ and 17 % of the maximum of the average measured shoot biomass $(max(\bar{\mathbf{y}}_{Shoot}))$. Where $\bar{\mathbf{y}}_{Shoot} = \bar{\mathbf{y}}_{Leaf} + \bar{\mathbf{y}}_{Stem} + \bar{\mathbf{y}}_{Pod}$. The root error term was also divided by N to give it equal weight as one time point in the leaf, stem or pod error terms.

$$\min_{\alpha, \beta, \mathbf{r_s}} E_{Leaf} + E_{Stem} + E_{Pod} + E_{Root} \tag{14}$$

$$E_i = \frac{1}{N} \sum_{j=1}^{N} w_{i,j} \left(\frac{\hat{y}_{i,j} - \bar{y}_{i,j}}{max(\bar{\mathbf{y}_i})} \right)^2 i \in \{\text{Leaf, Stem, Pod}\} \tag{15}$$

$$E_{Root} = \frac{1}{N} \left(\frac{max(\hat{\mathbf{y}}_{Root}) - 0.17 \cdot max(\bar{\mathbf{y}}_{Shoot})}{0.17 \cdot max(\bar{\mathbf{y}}_{Shoot})} \right)^2 \tag{16}$$

We further assigned a heavier cost to prediction errors at the time points where the measured leaf, stem and pod biomasses had smaller amounts of variation by defining weights, $w_{i,j}$, that were inversely proportional to the amount of variation in the experimental measurements (Equation (17)).

$$w_{i,j} = \ln \frac{1}{\sigma_{i,j} + \varepsilon} \ i \in \{\text{Leaf, Stem, Pod}\}, j = 1...N \tag{17}$$

$\sigma_{i,j}$ is the standard deviation of the measured biomasses of that plant organ at measured time j, and ε is a small value (e.g. 1e-5). We provide the code used to perform this parameter optimization in the GitHub repository (https://github.com/cropsinsilico/Soybean-BioCro).

2.3 Weather data and processing

The temperature, relative humidity, wind speed and photosynthetically active radiation (PAR) for the 2002, 2004–06 soybean growing season were obtained from the NOAA-ESRL SURFRAD Bondville, IL site which is ~7 miles west of the SoyFACE facility. The data can be accessed at ftp://aftp.cmdl.noaa.gov/data/radiation/surfrad/Bondville_IL/. The SURFRAD measurements were recorded at 3-min intervals. For each hour, we took the average of the reported measurements across that hour. The SURFRAD data come with quality check flags to indicate whether there are potential issues with the reported measurements. Any measurements that were flagged as having a quality issue were removed. Any missing measurements that remained after averaging the measurements over all the recorded points that hour were imputed from other data. We describe the data imputation steps in detail in the GitHub repository (https://github.com/cropsinsilico/Soybean-BioCro). No measurements were imputed for the 2002 soybean growing season. For the 2004 soybean growing season, 0.8 % of the PAR, and 0.3 % of the humidity, temperature and wind speed measurements were imputed. For the 2005 growing season, 0.03 % of the wind speed measurements were imputed. For the 2006 soybean

growing season, 0.3 % of the PAR, 0.8 % of the humidity, 0.3 % of the temperature and 3.9 % of the wind speed measurements were imputed.

Daily precipitation was obtained from the Illinois Water and Atmospheric Resources Monitoring (WARM) site in Champaign, IL which is 3 miles north of the SoyFACE facility. The daily precipitation was evenly divided across the 24 h in a day. Day length was calculated using an oscillator clock model included in BioCro (Lochocki and McGrath 2021). The code for producing the weather input files is provided in the linked GitHub repository.

3. RESULTS

3.1 Predicting soybean leaf, stem and pod biomasses in ambient and elevated $[CO_2]$

We used an evolutionary optimization algorithm (Mullen *et al.* 2011) to identify the parameters for the BioCro partitioning and senescence modules [see **Supporting Information—Table S5**] that minimized the predicted error of the leaf, stem and pod biomasses under the average ambient atmospheric $[CO_2]$ (372 ppm) of 2002 (Fig. 1A) and 2005 (Fig. 1B) growing seasons (Table 1, see Methods).

We simulated the 2002 and 2005 growing seasons under elevated atmospheric $[CO_2]$ (550 ppm) conditions, altering only the associated Catm (atmospheric $[CO_2]$) parameter. We further simulated the 2004 and 2006 growing seasons under both ambient and elevated $[CO_2]$ conditions (Fig. 2). For all four growing seasons, the model predicted leaf and stem biomasses that were consistent with the experimental observations under both ambient and elevated $[CO_2]$ conditions (Fig. 2A–G) and with similar MSEs as the 2002 and 2005 predictions under ambient $[CO_2]$ that were used for parameter fitting (Table 1).

Similarly, the predicted pod biomasses for 2002, 2005 and 2006 were also consistent with the experimental observations (Fig. 2I, K and L). For the 2004 growing season, however, the model over-predicted the maximum pod biomass by ~25 and ~15 % in ambient and elevated $[CO_2]$, respectively (Fig. 2J; Table 1).

Predicted leaf area index (LAI) measurements were consistent with experimental measurements under ambient and elevated $[CO_2]$ for the 2004 season and under elevated $[CO_2]$ for the 2006 growing season (Fig. 3A and C). The model under-predicted the LAI for the first half of the 2006 growing season under ambient $[CO_2]$ (Fig. 3C); however, this did not correspond to an under-prediction of the leaf biomass

Table 1. Mean square errors of simulated versus measured leaf, stem and pod biomasses. The rows highlighted in grey indicate the cases that were used for parameter fitting.

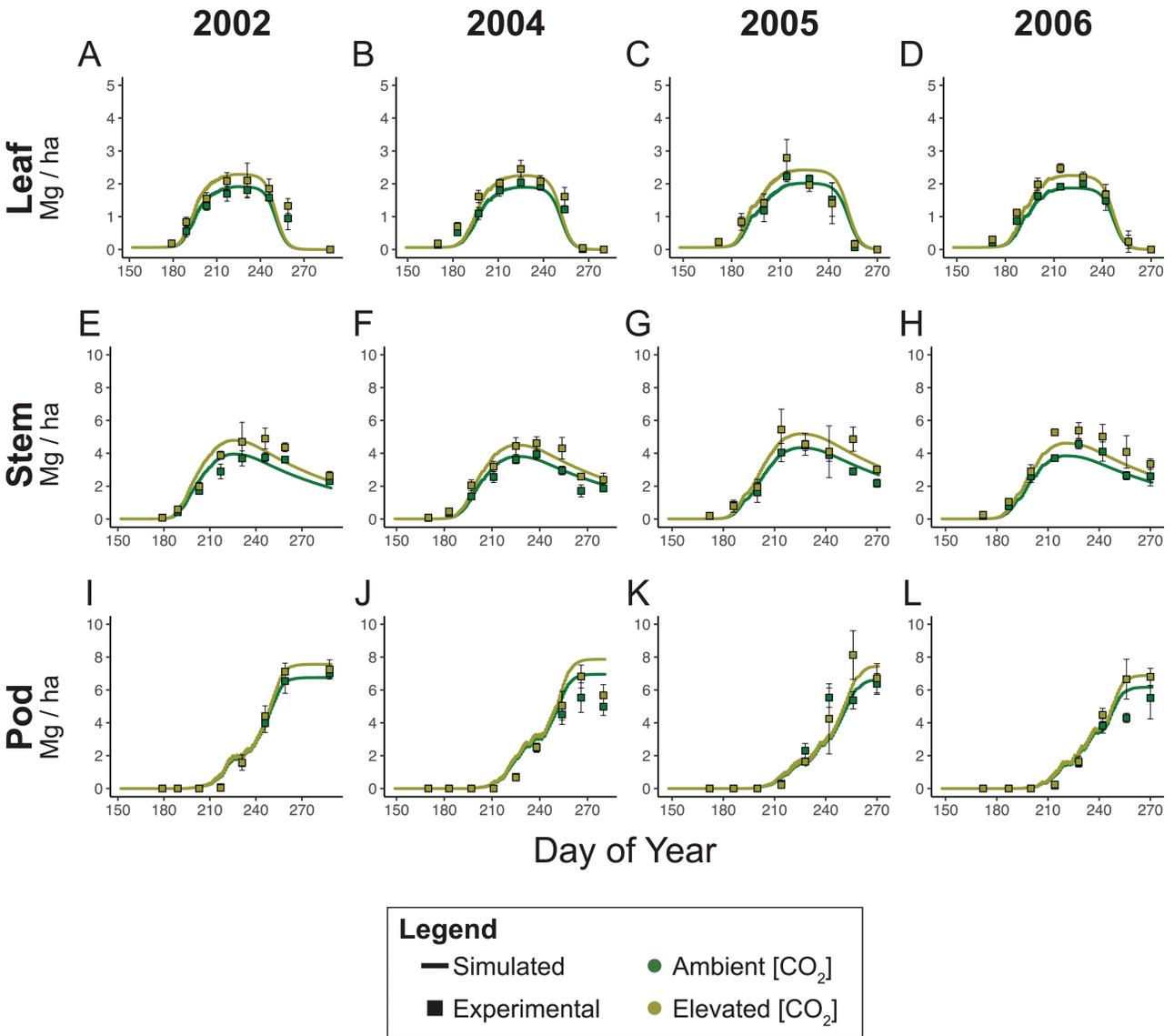| Year	$[CO_2]$ ppm	Leaf	Stem	Pod
2002	372 (fit)	0.29	0.49	0.30
	550	0.43	0.59	0.37
2004	372	0.24	0.39	0.96
	550	0.36	0.43	1.00
2005	372 (fit)	0.29	0.36	0.95
	550	0.44	0.58	0.80
2006	372	0.16	0.44	0.69
	550	0.22	0.69	0.40

Figure 2. Predicted and experimental (A–D) leaf, (E–H) stem and (I–L) pod biomasses over the 2002 and 2004–06 growing seasons under ambient (372 ppm) and elevated (550 ppm) atmospheric [CO_2].

(Fig. 2D). For the 2005 growing season LAI was over-predicted in the middle of the growing season for both ambient and elevated [CO_2] (Fig. 3B). The experimental LAI and leaf biomass measurements (Fig. 2C) indicate that in 2005, the crops started to senesce earlier than the predicted onset of senescence in the model.

For all of the four growing seasons, the model predicted that the maximum leaf, stem, shoot and pod biomasses were increased by ~20, ~20, ~14 and ~12 %, respectively, under elevated [CO_2] compared to ambient conditions (Table 2). Conversely, the predicted pod to shoot ratio, which is related to the harvest index, decreased by ~1.5 % (Table 2). These predictions are consistent with experimental results from soybean FACE experiments (Morgan *et al.* 2005; He *et al.* 2014; Bishop *et al.* 2015).

The over-predicted pod biomass in 2004 (Fig. 2J) indicates that there was some other environmental stressor impacting the pod biomass that was not accurately accounted for in our model. The recorded rainfall and predicted soil-water content were similar across the four growing seasons [see **Supporting Information—Fig. S1**], suggesting that drought was not the cause of the over-prediction. The average temperature during the 2004 growing season, 19.7 °C, however was 2–3 degrees cooler than the average temperatures of the 2002, 2005 and 2006 growing seasons, which were 22.3, 23.1 and 21.9°C, respectively (Fig. 4). BioCro's C_3 photosynthesis module (Eqs S32–S46) includes functions that describe the temperature-dependent response of the C_3 photosynthetic machinery at the biochemical and biophysical levels for Rubisco (Bernacchi *et al.* 2001), RuBP (Bernacchi *et al.* 2003)
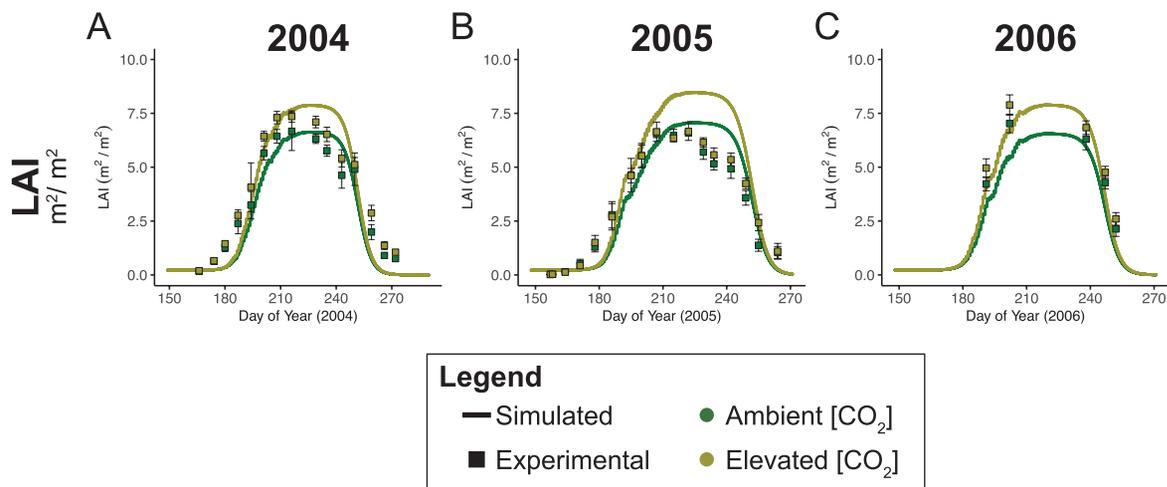
Figure 3. Predicted and measured LAI under ambient and elevated $[CO_2]$ for 2004–06.

Table 2. Relative maximum biomasses (elev. $[CO_2]$/amb. $[CO_2]$).

Year	Leaf	Stem	Shoot	Pod	Pod:Shoot
2002	1.21	1.21	1.14	1.12	0.98
2004	1.19	1.18	1.14	1.13	0.99
2005	1.20	1.19	1.14	1.13	0.99
2006	1.20	1.20	1.14	1.12	0.98

and TPU (Yang *et al.* 2016) limiting photosynthesis rates. During the late reproductive stages, where pod was over-predicted in 2004, the 2002, 2004 and 2006 growing seasons all saw similar cold temperatures, 17.2, 16.4, 16.6 °C on average, respectively. This suggests that cool temperatures during this period were not responsible for the discrepancy between the model prediction and the measured biomass. During the early reproductive stages, however, 2004 experienced colder temperatures than the other 3 years, with an average temperature of 20.0 °C for 2004 and 24.1, 24.1 and 24.0 °C in 2002, 2005 and 2006, respectively, during this period (Fig. 4). The predicted leaf and stem biomasses during this early reproductive phase of the 2004 growing season were consistent with experimental results (Fig. 2B and F). The onset of pod biomass was predicted to begin earlier than observed during this same period, but the predicted early onset does not completely account for the over-prediction (Fig. 2J). These results indicate that the temperature effects on photosynthesis do not account for the discrepancy between the predicted and measured pod biomass, but suggest that there may be additional temperature stress effects occurring in the early reproductive phase that limit the sink strength of the pod in the late reproductive stages.

4. DISCUSSION

Here, a soybean version of the semi-mechanistic BioCro crop growth model was successfully developed. This marks the first food crop to be included in the BioCro family of crop models. We added two new sets of modules to BioCro: (i) logistic-based carbon partitioning and senescence modules, and (ii) a development rate module that uses photothermal functions to calculate soybean development over a growing season. While these modules were specifically developed for Soybean-BioCro, they can also be used to model other BioCro crops. We parameterized the partitioning and senescence modules on leaf, stem and pod biomass data collected across two growing seasons, 2002 and 2005, at the SoyFACE facility in Urbana, IL under ambient atmospheric $[CO_2]$ conditions. Using these parameters, we successfully predicted the LAI and leaf, stem and pod biomasses in elevated $[CO_2]$ for 2002 and 2005, and ambient $[CO_2]$ and elevated $[CO_2]$ levels for the 2004 and 2006 growing seasons at the SoyFACE facility. In most cases, we were able to use BioCro to accurately predict soybean LAI and leaf, stem and pod biomasses over the growing seasons by only changing the input atmospheric $[CO_2]$ levels (Fig. 2). The exceptions were the predicted pod biomass during the 2004 growing season (Fig. 2J) and LAI for part of the 2005 and 2006 growing seasons (Fig. 3B and C). For the 2004 growing season, the BioCro model over-predicted the maximum pod biomass under both ambient and elevated $[CO_2]$. One explanation for this discrepancy could be the colder temperatures during the early reproductive stages of the 2004 season (Fig. 4B). Cooler temperatures, especially below 15 and 10 °C, during flowering and pod formation stages have been shown to impact pod formation, seed abortion and yield, resulting in fewer pods and seeds per pod (Musser *et al.* 1983; Seddigh *et al.* 1989; Srinivasan and Arihara 1994; Kurosaki and Yumoto 2003; Jähne *et al.* 2019). While the BioCro model includes functions to model the temperature response of the photosynthetic pathway, cold temperature stress has been shown to have a greater impact on carbon sinks (Wingler 2015; White *et al.* 2016; Sonnewald and Fernie 2018).

Experimental measurements of the leaf biomass from the 2005 growing season indicate an earlier onset of senescence than what was predicted by the model, especially under elevated $[CO_2]$ (Fig. 2C). While our senescence module is dependent on the soybean development rate, it does not incorporate other environmental feedbacks that may impact when a crop begins to senesce. Better
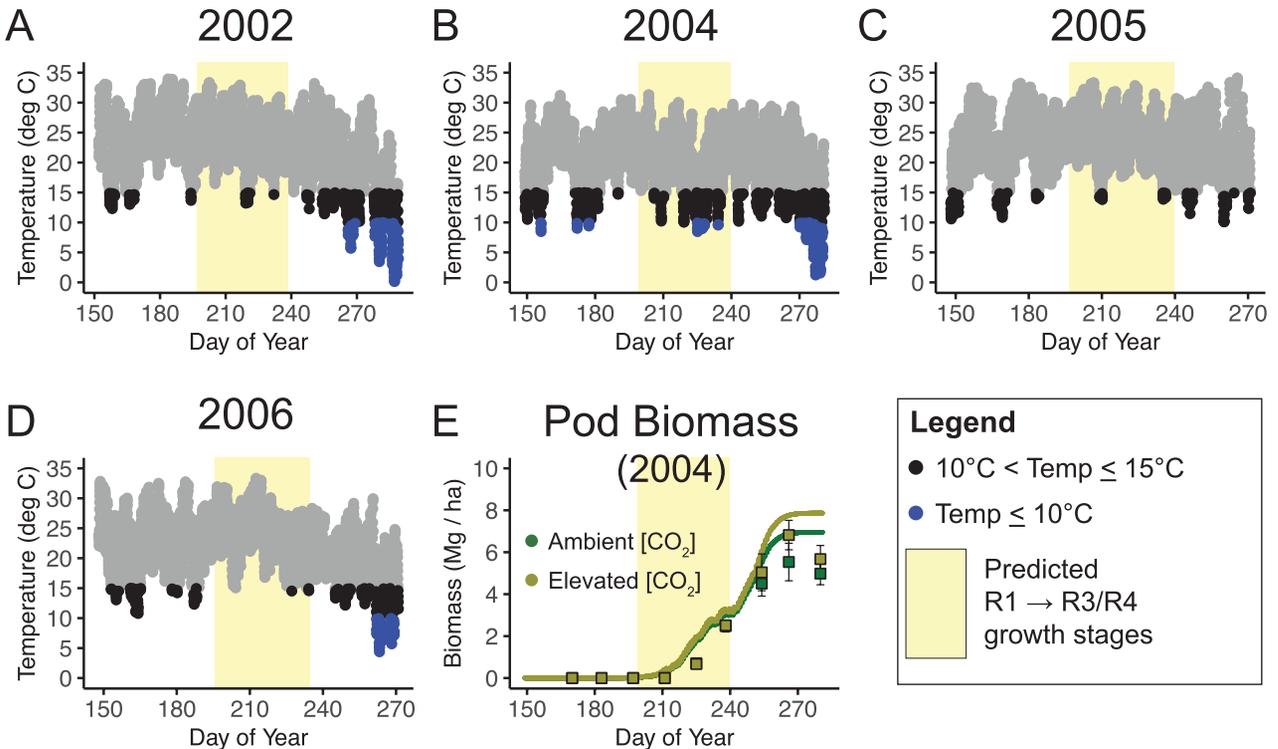
Figure 4. Hourly temperature measurements in Urbana, IL during the (A) 2002 and (B–D) 2004–06 soybean growing seasons. Temperatures ≤ 15 °C are in black and temperatures ≤10 °C are in blue. The predicted days when the soybean plants are in their R1–R3/R4 development stages are highlighted in yellow. (E) Predicted and experimental pod biomass in ambient and elevated [CO$_2$] during the 2004 growing season. This is the same plot as Fig. 2J, with the predicted range of the R1–R3/R4 development stages highlighted.

capturing sink constraints, and incorporating temperature or other environmental feedbacks to processes like carbon allocation and senescence, is an area for future improvement of the BioCro model. Overall, however, the results show the capability of Soybean-BioCro to predict beyond experience, particularly in the case of elevated [CO$_2$].

A strength of BioCro is its modularity. The BioCro framework was designed as a platform for multiple crops, where a crop species or cultivar is defined externally and not hardwired into the code. For example, the multilayer C$_3$ canopy module used in Soybean-BioCro can be used to model any C$_3$ crop without re-writing the module code since crop- or location-specific parameters are defined separately as inputs to a BioCro simulation. Further, new modules describing domain-specific processes can be easily added to the BioCro framework, and used with all crops parameterized in the framework. There has been recent effort towards developing multiscale crop models that span multiple levels of biological organization to explore crop adaptation and acclimation to a changing climate and genetic engineering strategies for improving crop productivity in current and future climates (Marshall-Colon *et al.* 2017; Peng *et al.* 2020; Matthews and Marshall-Colón 2021). Incorporating models of these behaviours will allow us to better study emergent properties like photosynthetic acclimation in future climate scenarios. Similar to how we developed new modules

for Soybean-BioCro, new modules can be developed that describe other primary and secondary metabolic processes or gene regulatory mechanisms. Using model coupling tools, such as the yggdrasil framework (Lang 2019), BioCro modules can also be developed that call models outside of the BioCro framework.

In conclusion, this study has shown that by incorporating the primary mechanism by which C$_3$ plants respond to rising [CO$_2$], Soybean-BioCro successfully predicted growth, partitioning and yield responses observed in a field soybean crop, providing a means to predict beyond experience. The modular nature of this semi-mechanistic model now provides a framework that can be extended to incorporate secondary effects of growth under future global change conditions including feedback effects on photosynthesis, partitioning and production (Kannan *et al.* 2019).

SUPPORTING INFORMATION

The following additional information is available in the online version of this article—

Text S1. The equations, parameters and initial values used in Soybean-BioCro.
Figure S1. Daily precipitation and predicted soil-water content for 2002, 2004–06.

ACKNOWLEDGEMENTS

SOURCES OF FUNDING

CONFLICT OF INTEREST

None declared.

CONTRIBUTIONS BY THE AUTHORS

M.L.M.: Conceptualization, methodology, software, formal analysis, writing—original draft, visualization; A.M.-C.: Conceptualization, writing—review & editing, supervision, funding acquisition; J.M.M.: Methodology, software, writing—review & editing; E.B.L.: Methodology, software, writing—review & editing; S.P.L.: Conceptualization, writing—review & editing, supervision.

DATA AVAILABILITY

The code used to reproduce the results and figures in this manuscript are available in a public GitHub repository: https://github.com/cropsinsilico/Soybean-BioCro.

LITERATURE CITED

Ainsworth EA, Rogers A, Leakey AD, Heady LE, Gibon Y, Stitt M, Schurr U. 2007. Does elevated atmospheric $[CO_2]$ alter diurnal C uptake and the balance of C and N metabolites in growing and fully expanded soybean leaves? *Journal of Experimental Botany* **58**:579–591.

Bernacchi CJ, Morgan PB, Ort DR, Long SP. 2005. The growth of soybean under free air $[CO(2)]$ enrichment (FACE) stimulates photosynthesis while decreasing *in vivo* Rubisco capacity. *Planta* **220**:434–446.

Bernacchi CJ, Pimentel C, Long SP. 2003. *In vivo* temperature response functions of parameters required to model RuBP-limited photosynthesis. *Plant, Cell & Environment* **26**:1419–1430.

Bernacchi CJ, Singsaas EL, Pimentel C, Portis AR Jr, Long SP. 2001. Improved temperature response functions for models of Rubisco-limited photosynthesis. *Plant, Cell & Environment* **24**:253–259.

Bishop KA, Betzelberger AM, Long SP, Ainsworth EA. 2015. Is there potential to adapt soybean (*Glycine max* Merr.) to future $[CO_2]$? An analysis of the yield response of 18 genotypes in free-air CO_2 enrichment. *Plant, Cell & Environment* **38**:1765–1774.

Choi IY, Hyten DL, Matukumalli LK, Song Q, Chaky JM, Quigley CV, Chase K, Lark KG, Reiter RS, Yoon MS, Hwang EY, Yi SI, Young ND, Shoemaker RC, van Tassell CP, Specht JE, Cregan PB. 2007. A soybean transcript map: gene distribution, haplotype and single-nucleotide polymorphism analysis. *Genetics* **176**:685–696.

Collins M, Knutti R, Arblaster J, Dufresne J-L, Fichefet T, Friedlingstein P, Gao X, Gutowski WJ, Johns T, Krinner G. Shongwe M, Tebaldi C, Weaver AJ, Wehner M. 2013. Long-term climate change: projections, commitments and irreversibility. In: Stocker TF, Qin D, Plattner G-K, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM, eds. *Climate change 2013— the physical science basis: contribution of working group i to the fifth assessment report of the Intergovernmental Panel on Climate Change*. Cambridge, UK and New York, NY: Cambridge University Press, 1029–1136.

Coutinho ID, Henning LMM, Döpp SA, Nepomuceno A, Moraes LAC, Marcolino-Gomes J, Richter C, Schwalbe H, Colnago LA. 2018. Flooded soybean metabolomic analysis reveals important primary and secondary metabolites involved in the hypoxia stress response and tolerance. *Environmental and Experimental Botany* **153**:176–187.

Das A, Eldakak M, Paudel B, Kim D-W, Hemmati H, Basu C, Rohila JS. 2016. Leaf proteome analysis reveals prospective drought and heat stress response mechanisms in soybean. BioMed Research International **2016**. Article ID: 6021047. https://doi.org/10.1155/2016/6021047.

Das A, Rushton PJ, Rohila JS. 2017. Metabolomic profiling of soybeans (*Glycine max* L.) reveals the importance of sugar and nitrogen metabolism under drought and heat stress. *Plants* **6**:21.

Haak DC, Fukao T, Grene R, Hua Z, Ivanov R, Perrella G, Li S. 2017. Multilevel regulation of abiotic stress responses in plants. *Frontiers in Plant Science* **8**:1564.

He Z, Xiong J, Kent AD, Deng Y, Xue K, Wang G, Wu L, Van Nostrand JD, Zhou J. 2014. Distinct responses of soil microbial communities to elevated CO_2 and O_3 in a soybean agro-ecosystem. *The ISME Journal* **8**:714–726.

Herring SC, Christidis N, Hoell A, Kossin JP, Schreck CJ, Stott PA. 2018. Explaining extreme events of 2016 from a climate perspective. *Bulletin of the American Meteorological Society* **99**:S1–S157.

Humphries SW. 2002. *Will mechanistically rich models provide us with new insights into the response of plant production to climate change?: development and experiments with WIMOVAC. PhD Thesis*, Department of Biological Sciences, University of Essex. https://eds.p.ebscohost.com/eds/detail/detail?vid=3&sid=89da26bdf66d-4fad-9b9f-f9eb766dd4a4%40redis&bdata=JnNpdGU9ZWRzLWxpdmU%3d#AN=uelc.b1535008&db=cat08605a.

Humphries SW, Long SP. 1995. WIMOVAC: a software package for modelling the dynamics of plant leaf and canopy photosynthesis. *Bioinformatics* **11**:361–371.

Hyten DL, Cannon SB, Song Q, Weeks N, Fickus EW, Shoemaker RC, Specht JE, Farmer AD, May GD, Cregan PB. 2010. High-throughput SNP discovery through deep resequencing of a reduced representation library to anchor and orient scaffolds in the soybean whole genome sequence. *BMC Genomics* **11**:38.

Jähne F, Balko C, Hahn V, Würschum T, Leiser WL. 2019. Cold stress tolerance of soybeans during flowering: QTL mapping and efficient selection strategies under controlled conditions. *Plant Breeding* **138**:708–720.

Jaiswal D, Souza APD, Larsen S, LeBauer DS, Miguez FE, Sparovek G, Bollero G, Buckeridge MS, Long SP. 2017. Brazilian sugarcane ethanol as an expandable green alternative to crude oil use. *Nature Climate Change* **7**:788–792.

Kannan K, Wang Y, Lang M, Challa GS, Long SP, Marshall-Colon A. 2019. Combining gene network, metabolic and leaf-level models shows means to future-proof soybean photosynthesis under rising CO_2. *In Silico Plants* **1**:diz008; doi:10.1093/insilicoplants/diz008.

Kurosaki H, Yumoto S. 2003. Effects of low temperature and shading during flowering on the yield components in soybeans. *Plant Production Science* **6**:17–23.

Lam HM, Xu X, Liu X, Chen W, Yang G, Wong FL, Li MW, He W, Qin N, Wang B, Li J, Jian M, Wang J, Shao G, Wang J, Sun SS, Zhang G. 2010. Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. *Nature Genetics* **42**:1053–1059.

Lang M. 2019. Yggdrasil: a Python package for integrating computational models across languages and scales. *In Silico Plants* **1**:diz001; doi:10.1093/insilicoplants/diz001.

Larsen S, Jaiswal D, Bentsen NS, Wang D, Long SP. 2016. Comparing predicted yield and yield stability of willow and Miscanthus across Denmark. *GCB Bioenergy* **8**:1061–1070.

Leakey ADB, Xu F, Gillespie KM, McGrath JM, Ainsworth EA, Ort DR. 2009. Genomic basis for stimulated respiration by plants growing under elevated carbon dioxide. *Proceedings of the National Academy of Sciences of the United States of America* **106**:3597–3602.

Lochocki EB, McGrath JM. 2021. Integrating oscillator-based circadian clocks with crop growth simulations. *In Silico Plants* **3**:diab016; doi:10.1093/insilicoplants/diab016.

Marshall-Colon A, Long SP, Allen DK, Allen G, Beard DA, Benes B, von Caemmerer S, Christensen AJ, Cox DJ, Hart JC, Hirst PM, Kannan K, Katz DS, Lynch JP, Millar AJ, Panneerselvam B, Price ND, Prusinkiewicz P, Raila D, Shekar RG, Shrivastava S, Shukla D, Srinivasan V, Stitt M, Turk MJ, Voit EO, Wang Y, Yin X, Zhu XG. 2017. Crops *in silico*: generating virtual crops using an integrative and multi-scale modeling platform. *Frontiers in Plant Science* **8**:786.

Matthews ML, Marshall-Colón A. 2021. Multiscale plant modeling: from genome to phenome and beyond. *Emerging Topics in Life Sciences* **5**:231–237.

Miguez FE, Maughan M, Bollero GA, Long SP. 2012. Modeling spatial and dynamic variation in growth, yield, and yield stability of the bioenergy crops *Miscanthus × giganteus* and *Panicum virgatum* across the conterminous United States. *GCB Bioenergy* **4**:509–520.

Miguez FE, Zhu X, Humphries S, Bollero GA, Long SP. 2009. A semimechanistic model predicting the growth and production of the bioenergy crop *Miscanthus × giganteus*: description, parameterization and validation. *GCB Bioenergy* **1**:282–296.

Morgan PB, Bollero GA, Nelson RL, Dohleman FG, Long SP. 2005. Smaller than predicted increase in aboveground net primary production and yield of field-grown soybean under fully open-air $[CO_2]$ elevation. *Global Change Biology* **11**:1856–1865.

Mullen KM, Ardia D, Gil DL, Windover D, Cline J. 2011. DEoptim: an R package for global optimization by differential evolution. *Journal of Statistical Software* **40**:1–26.

Muñoz N, Li MW, Ngai S-M, Lam H-M. 2016. 4—use of proteomics to evaluate soybean response under abiotic stresses. In: Miransari M, ed. *Abiotic and biotic stresses in soybean production*. San Diego, CA: Academic Press, 79–105.

Musser RL, Thomas SA, Kramer PJ. 1983. Short and long term effects of root and shoot chilling of ransom soybean. *Plant Physiology* **73**:778–783.

Nafziger E. 2016. Soybean: agronomy. In: Wrigley C, Corke H, Seetharaman K, Faubion J, eds. *Encyclopedia of food grains*, 2nd edn. Oxford: Academic Press, 251–256.

Ordóñez RA, Archontoulis SV, Martinez-Feria R, Hatfield JL, Wright EE, Castellano MJ. 2020. Root to shoot and carbon to nitrogen ratios of maize and soybean crops in the US Midwest. *European Journal of Agronomy* **120**:126130.

Osborne T, Gornall J, Hooker J, Williams K, Wiltshire A, Betts R, Wheeler T. 2015. JULES-crop: a parametrisation of crops in the Joint UK Land Environment Simulator. *Geoscientific Model Development* **8**:1139–1155.

Pagano MC, Miransari M. 2016. 1—the importance of soybean production worldwide. In: Miransari M, ed. *Abiotic and biotic stresses in soybean production*. San Diego, CA: Academic Press, 1–26.

Peng B, Guan K, Tang J, Ainsworth EA, Asseng S, Bernacchi CJ, Cooper M, Delucia EH, Elliott JW, Ewert F, Grant RF, Gustafson DI, Hammer GL, Jin Z, Jones JW, Kimm H, Lawrence DM, Li Y, Lombardozzi DL, Marshall-Colon A, Messina CD, Ort DR, Schnable JC, Vallejos CE, Wu A, Yin X, Zhou W. 2020. Towards a multiscale crop modelling framework for climate change adaptation assessment. *Nature Plants* **6**:338–348.

Prusinkiewicz P, Runions A. 2012. Computational models of plant development and form. *The New Phytologist* **193**:549–569.

Qiu LJ, Xing LL, Guo Y, Wang J, Jackson SA, Chang RZ. 2013. A platform for soybean molecular breeding: the utilization of core collections for food security. *Plant Molecular Biology* **83**:41–50.

Seddigh M, Jolliff GD, Orf JH. 1989. Night temperature effects on soybean phenology. *Crop Science* **29**:400–406.

Setiyono T, Weiss A, Specht J, Bastidas A, Cassman K, Dobermann A. 2007. Understanding and modeling the effect of temperature and daylength on soybean phenology under high-yield conditions. *Field Crops Research* **100**:257–271.

Song Q, Hyten DL, Jia G Quigley CV, Fickus EW, Nelson RL, Cregan PB. 2015. Fingerprinting soybean germplasm and its utility in genomic research. *G3: Genes, Genomes, Genetics* **5**:1999–2006.

Sonnewald U, Fernie AR. 2018. Next-generation strategies for understanding and influencing source–sink relations in crop plants. *Current Opinion in Plant Biology* **43**:63–70.

Srinivasan A, Arihara J. 1994. Soybean seed discoloration and cracking in response to low temperatures during early reproductive growth. *Crop Science* **34**:1611–1617.

Tans P, Keeling R. 2021. Global monitoring laboratory—carbon cycle greenhouse gases—trends in atmospheric carbon dioxide. https://www.esrl.noaa.gov/gmd/ccgg/trends/data.html. Accessed December 9 2021.

United Soybean Board. 2018. What are soybeans used for? [Internet]. https://www.unitedsoybean.org/hopper/what-are-soybeans-used-for/.

Wang D, Heckathorn SA, Barua D, Joshi P, Hamilton EW, Lacroix JJ. 2008. Effects of elevated CO_2 on the tolerance of photosynthesis to acute heat stress in C_3, C_4, and CAM species. *American Journal of Botany* **95**:165–176.

Wang D, Jaiswal D, LeBauer DS, Wertin TM, Bollero GA, Leakey AD, Long SP. 2015. A physiological and biophysical model of coppice willow (*Salix* spp.) production yields for the contiguous USA in current and future climate scenarios. *Plant, Cell & Environment* **38**:1850–1865.

White AC, Rogers A, Rees M, Osborne CP. 2016. How can we make plants grow faster? A source–sink perspective on growth rate. *Journal of Experimental Botany* **67**:31–45.

Wingler A. 2015. Comparison of signaling interactions determining annual and perennial plant growth in response to low temperature. *Frontiers in Plant Science* **5**:794.

Yang JT, Preiser AL, Li Z, Weise SE, Sharkey TD. 2016. Triose phosphate use limitation of photosynthesis: short-term and long-term effects. *Planta* **243**:687–698.

Zhou Z, Jiang Y, Wang Z, Gou Z, Lyu J, Li W, Yu Y, Shu L, Zhao Y, Ma Y, Fang C, Shen Y, Liu T, Li C, Li Q, Wu M, Wang M, Wu Y, Dong Y, Wan W, Wang X, Ding Z, Gao Y, Xiang H, Zhu B, Lee SH, Wang W, Tian Z. 2015. Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nature Biotechnology* **33**:408–414.