Developing an eco-physiological process-based model of soybean growth and yield (MATCRO-Soy v.1): Model calibration and evaluation

Astrid Yusara[1,8], Tomomichi Kato[2]*, Elizabeth A. Ainsworth[3*], Rafael Battisti[4], Etsushi Kumagai[5], Satoshi Nakano[6], Yushan Wu[7], Yutaka Tsutsumi-Morita[8], Kazuhiko Kobayashi[8], Yuji Masutomi[8]

[1]Graduate School of Agriculture, Hokkaido University, Sapporo, Hokkaido, Japan
[2]Research Faculty of Agriculture, Hokkaido University, Sapporo, Hokkaido, Japan
[3]Departments of Crop Sciences and Plant Biology, University of Illinois Urbana-Champaign, Urbana, Illinois, USA
[4]Escola de Agronomia, Universidade Federal de Goiás, Goiânia, Brazil
[5]Division of Agro-Environment Research Agricultural Meteorology Group, Tohoku Agricultural Research Center, NARO, Tsukuba, Ibaraki, Japan
[6]Central Region Agricultural Research Center, NARO, Tsukuba, Ibaraki, Japan
[7]College of Agronomy, Sichuan Agricultural University, Chengdu, PR China
[8]Center for Climate Change Adaptation, National Institute for Environmental Studies, Tsukuba, Ibaraki, Japan

Correspondence to: Tomomichi Kato (tkato@agr.hokudai.ac.jp)

Abstract. MATCRO-Soy is an eco-physiological process-based crop model for soybean (*Glycine max* L. (Merr.)). It was developed by modifying the parameters of MATCRO-Rice, which integrates crop growth processes with a land surface model. The original model was modified using data from the literature and field experiments conducted in countries around the world. The reliability of the model was extensively validated by comparing the simulated yields with observed yields at global, national, and grid-cell levels. Moderate correlations were detected between the yields predicted by MATCRO-Soy and yield data in the Food and Agriculture Organization's FAOSTAT database, with correlation coefficients of 0.81 ($p < 0.001$) for the global average yield and 0.45 ($p < 0.01$) for the global average detrended yield over a 34-year period (1981–2014). Furthermore, validation at the grid-cell level revealed a statistically significant correlation between the MATCRO-Soy simulated yield and the observed yield in 66 % of the grid cells in the global yield map. These results highlight the model's ability to reproduce soybean yield under different environmental conditions, integrating soil water availability and nitrogen fertilizer levels. The MATCRO-Soy model may enhance our understanding of crop physiology, especially crop responses to climate change. Its application may support efforts to reduce uncertainty in projections of the effects of climate change on soybean crops.

1 Introduction

Crop growth models are widely used for estimating yield, optimizing agricultural management practices, evaluating the effects of climate change, and informing decision-making about food security strategies (Adeboye et al., 2021; Cuddington et al., 2013; Hoogenboom, 2000). Given the significant impact of weather variability on global crop yields (Müller et al., 2017; Ray et al., 2015), process-based models can predict the effects of long-term climate change on productivity by accounting for the effects of key climatic factors on physiological processes that are represented in the model (Boote et al., 2013; Cuddington et al., 2013; Fodor et al., 2017; Jones et al., 2017; Marin et al., 2014; Stöckle and Kemanian, 2020). Process-based models explicitly incorporate the crucial eco-physiological processes of photosynthesis and stomatal conductance. Thus, the predictive ability of these models is improved under varying climate scenarios compared with that of models that focus on the empirical

relationship between absorbed radiation and assimilation through radiation use efficiency (Jin et al., 2018). Hence, crop models are useful for capturing the complexity of soil–crop–climate interactions for ensuring food security, optimizing yields, promoting sustainability, and planning adaptation strategies (García-Tejero et al., 2011). Global-scale simulations are crucial for enhancing these efforts because they reflect the interactions between physiological processes and environmental factors, thereby supporting adaptive management practices and strengthening agricultural resilience.

The Agricultural Model Intercomparison and Improvement Project (AgMIP) has examined the performance of global gridded crop models (GGCMs) in simulating the potential impact of climate change on crop yields (Müller et al., 2017). The AgMIP has demonstrated that the simulated impacts of environmental factors on crop yields using a GGCM generally align with measured values, and that a model ensemble reduces uncertainty (Elliott et al., 2015). However, yield changes under future climate change scenarios show inconsistent results and greater variability in soybean (*Glycine max* L. (Merr.)) than in other crops, because of model discrepancies (Jägermeyr et al. 2021). Despite being a major crop, soybean has been studied less extensively than other crops in terms of its response to changing environments (Ruane et al., 2017; Kothari et al., 2022). Therefore, the development of a new soybean model is needed to reduce uncertainties in climate change impact assessments.

It is important to use diverse types of crop models and ensure model diversity to understand the uncertainties of simulations, because relying on a single model can lead to biased results. To our knowledge, there are only five process-based models for global-scale soybean yield estimation with leaf-level photosynthesis and stomatal conductance parameters; namely LPJ-GUESS (Ma et al., 2022), LPJmL (Wirth et al., 2024), ORCHIDEE-crop (Wu et al., 2016), PRYSBI2 (Sakurai et al., 2014), and JULES (Leung et al., 2020). Simulations for soybean using process-based models are relatively uncommon. Thus, further development and validation of process-based models that incorporate leaf-level photosynthesis and stomatal conductance parameters are essential.

MATCRO-Rice (Masutomi et al., 2016a,b) is an ecosystem process-based model for crops embedded into the land surface model of Minimal Advanced Treatments of Surface Interaction and Runoff (MATSIRO; Takata et al., 2003). The crop growth model is further explained in Section 2. MATCRO-Rice uses state variables to exchange information (e.g., temperature, soil moisture, transpiration, leaf area index, and photosynthesis rate) between the land surface model and crop growth model. The MATCRO-Rice model incorporates mechanisms related to photosynthesis and stomatal conductance to assess the impact of greenhouse gases on carbon and water fluxes on crop yield. Masutomi et al. (2019) described the implementation of ozone effects as one of these mechanisms, highlighting the model's capability to account for environmental stressors. Furthermore, MATCRO-Rice has been applied at the regional scale, and it has been used to measure climate impacts, which are important for developing adaptation strategies (Kinose and Masutomi, 2020; Masutomi, et al., 2016b).

Here, we developed a new process-based model for soybean, MATCRO-Soy v.1, which incorporates diverse biological processes and environmental interactions that drive plant growth and adaptation to changing conditions. Adapted from MATCRO-Rice, the new model was applied to soybean by parameterizing key processes using experimental data and findings from the literature. The current version of MATCRO-Soy (v.1) was evaluated in a global-scale simulation, following a calibration process that considered essential photosynthesis mechanisms. This paper presents the model description in Sect. 2, the calibration process in Sect. 3, and the model evaluation in Sects. 4 and 5.

2 Model description

MATCRO-Soy is based on MATCRO-Rice, a process-based model for rice growth and yield. Here, the MATCRO-Rice model has been modified for use in soybean. MATCRO-Rice is a combined land surface and crop growth model used to explore the land–atmosphere interaction in rice fields. Unlike MATCRO-Rice v.1, MATCRO-Soy focuses on yield simulation and omits the calculation of sensible and latent heat fluxes in the energy balance to reduce computational complexity while maintaining accuracy in simulating soybean growth and yield.

2.1 Overview of MATCRO-Soy

MATCRO-Soy includes three main modules: phenology, photosynthesis, and carbon partitioning (Figure 1). The phenology module simulates crop phenological development over time based on heat unit accumulation. The module directs the progression of carbon assimilation and partitioning by monitoring plant developmental stages from sowing to harvest. The phenology module simulates developmental stages based on the developmental rate from sowing to harvest. The developmental stage influences key processes such as glucose production and allocation across plant organs. The photosynthesis module initially estimates gross primary production (GPP) and respiration at the leaf level using the Farquhar model (Farquhar et al., 1980), and then extends the estimation of net primary production (NPP) to the canopy level, following the concept introduced by de Pury and Farquhar (1997). It considers the electron-transport-limited rate of photosynthesis, Rubisco-limited photosynthesis, and leaf respiration to estimate NPP at the leaf level.

The photosynthesis and carbon partitioning modules are closely linked, because carbon assimilated from photosynthesis is subsequently allocated to different plant organs. The NPP is stored in glucose and starch reserves. The carbon partitioning module distributes glucose to different organs (i.e., leaf, stem, root, and storage organ) using a method derived from the school of de Wit, which simulates biosynthetic processes (de Vries et al., 1989). It also accounts for leaf senescence, which influences nutrient cycling, crop productivity, and the leaf area index, thereby affecting canopy photosynthesis. Leaf senescence is simulated as a function of crop developmental stage, as defined by the phenology module. MATCRO incorporates the amount of nitrogen per leaf area (specific leaf nitrogen, *SLN*) as a key determinant of photosynthetic capacity. Root depth can indirectly affect photosynthesis because it influences the plant's ability to access water and nutrients from soil layers, further influencing plant growth within the model framework.

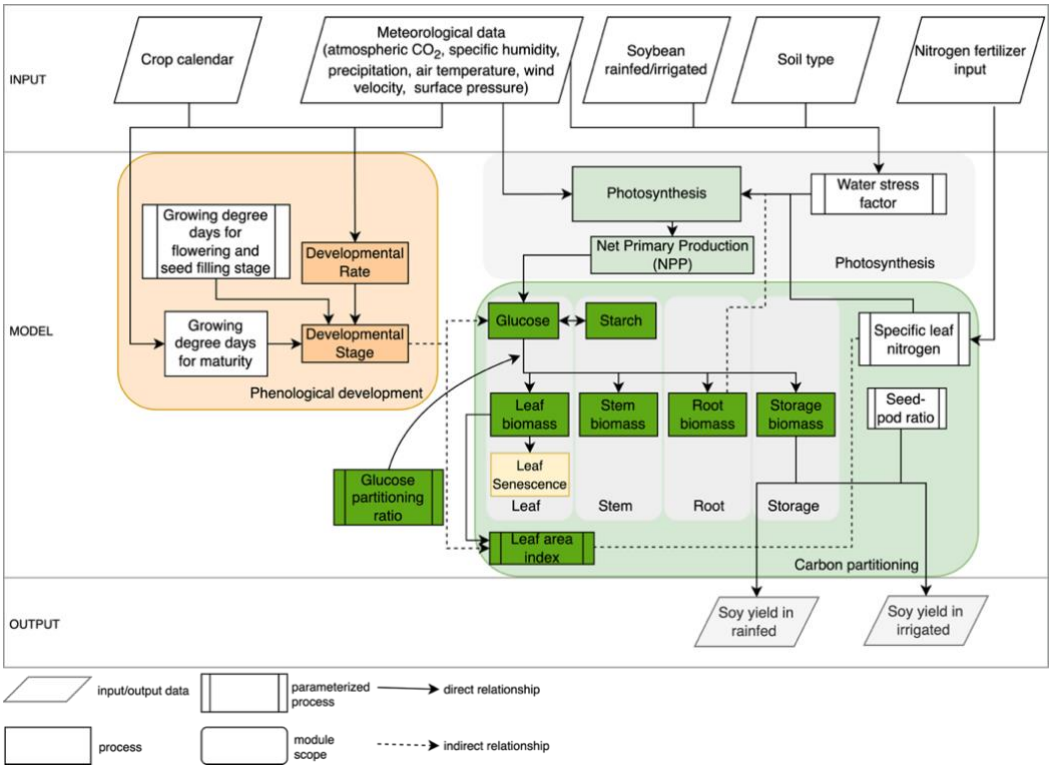

Figure 1. Flowchart diagram of soybean yield simulation by the MATCRO-Soy model.

The input data consisted of environmental variables obtained from meteorological forcings, soil type classifications, nitrogen fertilizer applications, and agricultural management practices such as irrigation and seed sowing. These inputs were crucial for setting the initial conditions and boundary parameters for the simulations. The output of MATCRO is crop yield (kg ha^{-1}) estimated for both irrigated and rainfed conditions on the basis of soil–crop interactions. First, we processed the parameterized growing degree days to maturity using crop calendar data to estimate the harvest time in the phenology module (see Sect. 2.2). The photosynthesis module includes limiting factors such as nitrogen fertilization and water stress, as detailed in Sect. 2.3. Then, crop growth is calculated based on developmental stage (Sect. 2.4). We conducted a parameterization process including phenological development, carbon partitioning, and photosynthesis limited by water stress and nitrogen uptake. The crop yield was estimated using the parameterized seed:pod ratio (see Sect. 2.5). The adjusted parameters in MATCRO-Soy are described in Sect. 2.6, where the key dynamic variables were parameterized over time to ensure a reliable estimate of carbon assimilation in soybean. This comprehensive approach allows MATCRO to account for the complex interactions among environmental conditions, crop physiology, and management practices, providing a robust framework for predicting crop yields and assessing agricultural productivity.

2.2 Crop phenological development

Phenological development refers to the timing of developmental events in response to environmental inputs. MATCRO calculates crop developmental stage (DVS) using an index ranging from 0 to 1, where DVS=0 is the sowing time and DVS=1 is maturity. This index is based on the integral of the temperature required to exceed the phenological changes. The module uses a formulation based on Bouman et al. (2001) as outlined in Eqs. (1)–(4) as follows:

$$DVS_t = GDD_t/GDD_m \tag{1}$$

$$GDD_t = \int_0^t DVR dt' \tag{2}$$

$$GDD_m = \int_0^m DVR dt' \tag{3}$$

$$DVR_t = \begin{cases} 0, & T_t < T_b \mid T_t > T_h \\ T_t - T_b, & T_b < T_t < T_o \\ \frac{(T_o - T_b)(T_h - T_t)}{(T_h - T_o)}, & T_o < T_t < T_h \end{cases} \tag{4}$$

where GDD_t and GDD_m indicate the growing degree days (°C days) used to estimate the development of plants during the growing season at time t and at maturity (m), respectively; DVR represents the developmental rate at t; and T_t represents the temperature at t. The parameters T_b, T_o, and T_h (°C) are crop-specific and represent the base, optimum, and highest temperatures for crop development, respectively.

The impact of temperature on phenological stage can differ among crop stages, as Boote et al. (1998) observed that cardinal temperatures (T_b, T_h, T_o) may differ between vegetative and reproductive stages. We followed de Vries et al. (1989) for cardinal temperatures during the growing season. This study parameterized the developmental stages as flowering (DVS_f), seed-filling (DVS_s), and maturation (DVS_m) on the basis of mean values calculated from the available observations for each stage (listed in Table 2). Calculations for each stage were based only on experiments where corresponding data were available. MATCRO uses these DVS parameters to define the period of leaf dry weight loss due to leaf senescence and the remobilization of starch reserves from the stem (Masutomi et al. 2016a). It was assumed that the corresponding phenological times in soybean are the middle of the flowering stage and the seed-filling stage, because leaf loss starts within those periods.

2.3 Carbon assimilation process

In the photosynthesis module of MATCRO-Soy, carbon assimilation is based on canopy photosynthesis, which is estimated from leaf-level photosynthesis calculated in sunlit and shaded conditions (Dai et al., 2004). The calculation includes the stomatal conductance response to relative humidity (Collatz et al., 1991). The net carbon assimilation (A_n) in MATCRO is calculated using the Farquhar model as further described in Masutomi et al. (2016a), expressed in Eq. (5) as given as

$$A_n = f(PAR, CO_{2leaf}, V_{\max}, \mathrm{P_a}, RH, T_{leaf}, f_w, BB_a, BB_b) \quad (5)$$

where A_n (mol(CO_2) m^{-2} s^{-1}) represents net carbon assimilation contributing to NPP for biomass growth. It is a function of the intensity of absorbed photosynthetic active radiation (PAR, in mol(photon) m^{-2} s^{-1}), the CO_2 concentration in the substomatal chamber (CO_{2leaf}, in Pa(CO_2) Pa(Air)$^{-1}$), maximum Rubisco capacity per unit leaf area ($V_{\max}$, in mol(CO_2) m^{-2} s^{-1}), air pressure (P_a, in Pa), relative humidity (RH), leaf temperature (T_{leaf}, in K), water stress factor (f_w, dimensionless), the slope (BB_a, in mol(H_2O)$m^{-2}s^{-1}$) and intercept (BB_b, in mol(H_2O) m^{-2} s^{-1}) of the Ball-Berry model of the relationship between crop assimilation and stomatal conductance per unit leaf area, relative humidity at the leaf surface, and ambient CO_2 concentration (Ball, 1988). The leaf temperature is assumed to be the same as the air temperature to simplify the calculation.

Rubisco activity is a key variable used to assess the rate of carbon entry into the photosynthetic pathway, because Rubisco catalyzes the crucial initial step of RuBP (ribulose-1,5-bisphosphate) carboxylation in photosynthetic carbon assimilation in C_3 plants (Sage, 2002; Xu et al., 2022). In MATCRO, V_{max} in Eq. (5) is calculated as follows:

$$V_{max} = V_{mc} f_w \quad (6)$$

$$V_{mc} = \bar{V}_{mc} \exp(\mathrm{c} - \Delta\mathrm{H}/\mathrm{R}T_{leaf}) \quad (7)$$

$$\bar{V}_{mc} = \frac{\int_0^{LAI} V_{cmax}(LAI)\,dLAI}{LAI} \quad (8)$$

where V_{max} and V_{mc} are, respectively, the maximum Rubisco capacity per unit leaf area with and without the water stress factor (f_w); V_{mc} is determined with a generic temperature response as described by Bernacchi et al. (2001); c and ΔH represent a scaling constant (c = 26.35) and activation energy (ΔH = 65.33 kJ mol^{-1}) of Rubisco's activity response to temperature changes; R is the molar gas constant in kJ mol^{-1}; $\bar{V}_{mc}$ is the maximum Rubisco capacity averaged over the canopy; and $V_{cmax}(LAI)$ denotes the vertical distribution of the maximum Rubisco capacity through the canopy, as determined by the vertical nitrogen distribution (see Eqs. (14) and (15)). The water stress factor, f_w, is determined based on the root distribution function ($f_r(i)$) multiplied by the water stress function at each soil layer ($f_{wstress,t}(i)$). The results are then summed across five soil layers (depths of 0.05, 0.25, 1, 2, and 4 m below the ground surface), as given in Eqs. (9)–(13) as follows:

$$f_w = \sum_{i=1}^{5} f_r(i) f_{wstress,t}(i) \quad (9)$$

$$f_r(i) = \left(\frac{3}{2}\right) \frac{({z_r}^2 - i^2)}{{z_r}^3} \quad (10)$$

$$z_r = \min\{z_{rootmax}, r_{root}(t - t_e)\} \quad (11)$$

where $f_r(i)$ is the distribution of roots in soil layer i (with value of 1–5). Root depth (z_r, in m) is calculated based on the root growth rate (r_{root}, mm day^{-1}) in timestep t (day) after the time of emergence (t_e, in d), and is limited by the maximum root

depth ($z_{rootmax}$, in m). t_e is assumed to be in the early developmental stage (0.012 of the growing period). The function $f_{wstress,t}$ represents a simplified version of the relationship between the soybean transpiration ratio and transpirable soil water devised by Ray and Sinclair (1998), given in Eq. (12).

$$f_{wstress,t}(i) = \begin{cases} \frac{1}{0.5}\, FAW(i),\ if\ FAW(i) \le 0.5 \\ 1,\ if\ FAW(i) > 0.5 \end{cases} \quad (12)$$

The value of the water stress function at timestep *t* ($f_{wstress,t}$) depends on soil water availability at soil layer *i* (FAW_i), which is the estimated soil water content based on the water flux between the soil layers during crop growth calculated by

$$FAW(i) = \frac{WSL(i) - WSL_{wilt}}{WSL_{FC} - WSL_{wilt}} \quad (13)$$

where $WSL(i)$, WSL_{wilt}, and WSL_{FC} represent the water levels in the soil layer *i*, at wilting point, and at field capacity, respectively. A value of $f_{wstress}$ equal to 1 indicates no water stress because the fraction of available soil water is adequate for crop growth.

$V_{cmax}(LAI)$ is the reference value for maximum Rubisco activity within the canopy (mol(CO_2) m^{-2} s^{-1}) at leaf area index (LAI, in m^2 m^{-2}) depth, limited by the exponential value of vertical distribution of leaf nitrogen (K_n), and the reference value for maximum Rubisco activity at the top of canopy (V_{ctop}, in mol(CO_2) m^{-2} s^{-1}), calculated as follows:

$$V_{cmax}(\mathrm{LAI}) = V_{ctop} \exp(-K_n LAI) \quad (14)$$

$$V_{ctop} = \max\left(aSLN^2 + bSLN + c, V_{ctopmax}\right) \quad (15)$$

For soybean, the V_{ctop} photosynthetic rate limited by the SLN is based on the relationship between Rubisco activity and leaf nitrogen content, as determined from experiments on soybean at the reproductive stage, summarized by Ainsworth et al. (2014), and for soybean at the reproductive stage, summarized by Qiang et al. (2022). This relationship is empirically represented by a polynomial quadratic equation limited by the maximum value of Rubisco activity at the top canopy ($V_{ctopmax}$ in mol(CO_2) m^{-2} s^{-1}). a, b, c are the quadratic coefficient, linear coefficient, and constant, respectively, from the relationship between the two variables based on data digitized using WebPlotDigitizer (Rohatgi, 2023).

MATCRO considers nitrogen fertilization input denoted as N_{fert} (unit: kg(N) ha^{-1}). This influences the amount of SLN (g(N) m^{-2}), particularly under conditions of limited nitrogen availability (La Menza et al., 2023; Thies et al., 1995). The *SLN* is determined by nitrogen supply (including biological nitrogen fixation, soil mineral nitrogen, and nitrogen fertilizer) and by plant demand. In MATCRO-Soy, the changes in *SLN* over the growing period are represented by a function derived from La Menza et al. (2023), who observed *SLN* under wide range of low- and high-nitrogen fertilization conditions (see Supplementary file Fig. S1). This function adjusts the *SLN* value during the crop growth period, and higher nitrogen fertilization levels result in a higher leaf nitrogen content. In the absence of empirical data for initial growth stages, the model assumes a gradual increase in nitrogen content. The simulated *SLN* under different nitrogen fertilization treatments is defined by

$$SLN = \begin{cases} SLN_{Y0} + \frac{(SLN_{Y1} - SLN_{Y0})(DVS - SLN_{X1})}{SLN_{X1}}, \ if\ DVS < SLN_{X1} \\ SLN_{Y2} + \frac{(SLN_{Y2} - SLN_{Y1})(DVS - DVS_f)}{(DVS_f - SLN_{X1})}, \ if\ SLN_{X1} \leq DVS < DVS_f \\ Y + \frac{(Y - SLN_{Y2})(DVS - DVS_s)}{(DVS_s - DVS_f)}, \ if\ DVS_f \leq DVS < DVS_s \\ SLN_{Y0} + \frac{(SLN_{Y0} - Y)(DVS - DVS_m)}{(DVS_m - DVS_s)}, \ if\ DVS_s \leq DVS \leq DVS_m \end{cases} \quad (16)$$

$$Y = SLN_{Y3,l} + \frac{SLN_{Y3,h} - SLN_{Y3,l}}{N_{fert,high}} * N_{fert} \quad (17)$$

SLN values vary across different phenological stages, as the developmental stage (DVS) of soybean plants progresses from 0 (at sowing) to 1 (at harvest). DVS_f, DVS_s, DVS_m, and SLN_{X1} are defined as the start of flowering, seed-filling, and maturity stages, and the point where the SLN pattern starts to change; with parameterized values of 0.4, 0.659, 1, and 0.15, respectively. SLN_{Y0}, SLN_{Y1}, SLN_{Y2}, $SLN_{Y3,l}$, and $SLN_{Y3,h}$ represent SLN at the initial stage, early decline, pre-flowering increase, subsequent decline phases during the reproductive stage under no (l) and high (h) nitrogen inputs with values of 0.75, 2.25, 1.7, 0.75, and 1.8, respectively. $N_{fert,high}$ refers to the high nitrogen fertilizer input used in the model for parameterization, as described in Table 2. Y is the observed gap in *SLN* between high- and low-nitrogen fertilizer treatments (g(N) m^{-2}) (see Supplementary file Fig. S1).

The growth stages were parameterized based on experimental datasets and align with those reported by Irmak et al. (2013), based on the growth stage classifications of Fehr and Caviness (1977). SLN primarily depends on nitrogen derived from biological fixation and soil nitrogen, either from natural sources or applied fertilizers. Nitrogen uptake, including biological nitrogen fixation and uptake from soil nitrogen, is implicitly captured through SLN that influences V_{cmax} [Eqs. (14) and (15)], and *SLN* as affected by applied fertilizers [Eqs. (16) and (17)].

2.4 Crop growth dynamics

The products of photosynthesis contribute to glucose reserves, which provide energy for growth during various developmental stages. The crop growth dynamics include a carbon biomass partitioning module to calculate the dry weight of each soybean organ (W_{organ} in kg ha^{-1}; see Eq. 18). This variable is the cumulative growth rate of dry weight (G_{organ} in kg ha^{-1} s^{-1}) during the time from emergence to harvest. Further details on this module can be found in Masutomi et al. (2016a).

$$W_{organ} = f(G_{organ}) \quad (18)$$

The W_{organ} is calculated separately for each soybean organ (i.e., leaf, stem, and pod including the seed, glucose reserves, and starch). The growth rate of dry weight (G_{organ} in kg ha^{-1} s^{-1}) is calculated based on the parameters of conversion factor of dry weight from glucose to organ ($F_{glu-organ}$ in kg ha^{-1}(kg ha^{-1})$^{-1}$) for leaf, stem, pod, root, and starch (listed in Table 1), and the ratio of glucose partitioned to organ (P_{organ}) for shoot, leaf, and pod (listed in Table 2). Shoot refers to aboveground biomass parts including the stem, leaf, and pod. G_{organ} is calculated for each organ and storage fraction (glucose, leaf, stem, pod, root, and starch) as described by:

$$G_{glu} = f(W_{leaf}, A_{glu}, R_{glu}) \quad (19)$$

$$G_{leaf} = G_{glu} P_{shoot} P_{leaf} F_{glu\text{-}leaf} \quad (20)$$

$$G_{stem} = G_{glu}P_{shoot}\left(P_{leaf} - P_{pod}\right) \times (1 - f_{starch})F_{glu-stem} \quad (21)$$

$$G_{pod} = G_{glu}P_{shoot}P_{pod}F_{glu-pod} \quad (22)$$

$$G_{root} = G_{glu}(1 - P_{shoot})F_{glu-root} \quad (23)$$

$$G_{starch} = G_{glu}P_{shoot}\left(P_{leaf} - P_{pod}\right)f_{starch}F_{glu-starch} \quad (24)$$

where G_{glu} (kg ha^{-1} s^{-1}) is the amount of glucose partitioned to soybean organs and reserves derived from a function of leaf dry weight (W_{leaf} in kg ha^{-1}), net carbon assimilation in glucose form (A_{glu} in kg(CH_2O) ha^{-1} s^{-1}), and glucose remobilized from starch reserves in the stem (R_{glu} in kg ha^{-1} s^{-1}); A_{glu} is A_n that has been already converted from CO_2 to glucose using the conversion factor 1.08×10^6 [kg ha^{-1} h^{-1}(mol m^{-2} s^{-1})$^{-1}$], which is the physical and chemical constant for the conversion; and R_{glu} is glucose remobilized from starch reserves in the stem, calculated using the ratio of the remobilization value. The R_{glu} is subtracted from the dry weight of starch reserves (W_{starch}). f_{starch} [kg ha^{-1}(kg ha^{-1})$^{-1}$] is the fraction of glucose allocated to starch reserves, calculated as stem dry weight loss.

As shown in Eqs. 20–24, G_{organ} was calculated based on the conversion factor of dry weight ($F_{glu-organ}$) and ratio of glucose partitioned to that organ (P_{organ}). The calculations for P_{organ} are shown in Eqs. (25)–(27):

$$P_{shoot} = \begin{cases} 1 - P_{root}, & if\ DVS = 0 \\ \frac{1-P_{root}(DVS_m - DVS)}{DVS_m}, & if\ 0 < DVS < DVS_m \\ 1, & if\ DVS \geq DVS_m \end{cases} \quad (25)$$

$$P_{leaf} = \begin{cases} P_{leaf0} + \frac{DVS}{DVS_{leaf1}}(P_{leaf1} - P_{leaf0}), & if\ DVS < DVS_{leaf1} \\ P_{leaf2} - \frac{(P_{leaf2} - P_{leaf1})}{DVS_{leaf2} - DVS_{leaf1}}(DVS_{leaf2} - DVS), & if\ DVS_{leaf1} \leq DVS < DVS_{leaf2} \\ 0, & if\ DVS \geq DVS_{leaf2} \end{cases} \quad (26)$$

$$P_{pod} = \begin{cases} 0, & if\ DVS < DVS_{pod1} \\ \frac{DVS - DVS_{pod1}}{DVS_{pod2} - DVS_{pod1}}, & if\ DVS_{pod1} \leq DVS < DVS_{pod2} \\ 1, & if\ DVS \geq DVS_{pod2} \end{cases} \quad (27)$$

P_{leaf0}, P_{leaf1}, P_{leaf2} represent the leaf:shoot glucose partitioning ratio when leaf growth first starts to decline (leaf0), leaves stop growing (leaf1), and at maturity (leaf2), respectively. DVS_{pod1} and DVS_{pod2} indicate the *DVS* values at which the pod:shoot glucose partitioning ratio begins to increase and eventually saturates, respectively (Fig. 2). Figure 2 in Sect. 3.2 visually represents the glucose partitioning ratio during crop growth as calibrated in this study.

During the calibration process, the glucose partitioned to each organ was adjusted for each developmental stage on the basis of experimental data, as further described in Sect. 3. However, in this module, the leaf dry weight decreases because of senescence. This is calculated as the loss of leaf dry weight (L_{leaf} in kg ha^{-1} s^{-1}) derived from the calibration of the ratio of glucose partitioned into the dead leaf (P_{dleaf} in s^{-1}), as outlined in Eqs. (28) and (29).

$$L_{leaf} = \begin{cases} 0 & , \ if\ DVS < DVS_{deadleaf1} \\ P_{dleaf}(W_{leaf} - W_{glu}), & if\ DVS \geq DVS_{deadleaf1} \end{cases} \quad (28)$$

$$P_{dleaf} = P_{deadleaf2} \frac{(DVS - DVS_{deadleaf1})}{(1 - DVS_{dealeaf1})} \quad (29)$$

The leaf area index (*LAI*) represents the leaf surface area relative to the ground area, calculated using Eq. 30. It directly influences the plant's ability to intercept solar radiation for photosynthesis.

$$LAI = \frac{W_{leaf} + W_{glu}}{SLW} \quad (30)$$

LAI is calculated from the estimated leaf dry weight (W_{leaf}, in kg ha^{-1}) and glucose reserves in leaves (W_{glu}, in kg ha^{-1}) divided by the specific leaf weight (SLW, in kg ha^{-1}). Glucose reserves are added to the leaf dry weight as a buffer, and affect leaf growth by storing carbohydrates that are not immediately required. *SLW* is the leaf dry weight per unit leaf area. The value of *SLW* dynamically changes during development according to the following exponential relationship:

$$SLW = SLW_{max} + (SLW_{min} - SLW_{max})\, exp(-SLW_x DVS) \quad (31)$$

where SLW_{max}, SLW_{min}, and SLW_x represent the maximum, minimum, and slope parameters, respectively, that define the values observed in the exponential relationship based on the experimental dataset summarized in Table 3.

2.5 Soybean yield estimation

Soybean yield is calculated from the pod dry weight at harvest ($W_{podharvest}$, in kg ha^{-1}) multiplied by the seed ratio parameter (SR), as given in Eq. (32).

$$Yield = W_{podharvest} \times SR \quad (32)$$

SR was derived from experimental datasets summarized in Table 3 and represents the ratio of yield (seed, kg ha^{-1}) to $W_{podharvest}$ at harvest time.

2.6 Soybean-specific parameters

MATCRO-Soy shares several parameters with MATCRO-Rice because both plants are C_3 species. However, unlike cereal crops, soybean plants can fix nitrogen. This characteristic is represented through changes in *SLN* during crop growth, as described in Eqs. (16) and (17). The crop-specific parameters reflect the unique physiological and chemical processes involved in soybean growth, but still align with the general framework of MATCRO-Rice. Key parameter adjustments are outlined in Table 1, where MATCRO employs a set of specific parameters to simulate crop growth and yield. These parameters include factors related to carbon allocation, root growth characteristics, and crop development based on cardinal temperatures. By accurately representing the unique physiological and biochemical characteristics of soybean plants, these parameters improve the precision of the model in predicting soybean yield.

MATCRO-Soy is intended for use in global-scale simulations; hence, it uses a single global parameterization as a standardized set of parameters applied worldwide. It uses a unified approach for modelling crop behaviour across different regions. It is assumed that the parameter values from the different treatments and cultivars are independent. Table 2 lists

variables parameterized within the model, including glucose partitioning, nitrogen, and photosynthetic capacity variables. Through the parameterization of these variables, the model can be adapted for various growing conditions and employed to assess the sensitivity of crop performance to different factors. These parameters are commonly used to evaluate the crop model's sensitivity to environmental changes and require further fine-tuning, as highlighted by simulations using other crop models (Battisti et al., 2018a).

Table 1. Crop-specific parameters used for MATCRO-Soy.

Parameters	Description	Value	Units	Source	Eq.
$F_{glu-leaf}$	conversion factor of dry weight from glucose to leaf	0.871	kg ha^{-1} (kg ha^{-1})$^{-1}$	de Vries et al. (1989)	(20)
$F_{glu-stem}$	conversion factor of dry weight from glucose to stem	0.810	kg ha^{-1} (kg ha^{-1})$^{-1}$	de Vries et al. (1989)	(21)
$F_{glu-pod}$	conversion factor of dry weight from glucose to pod	0.759	kg ha^{-1} (kg ha^{-1})$^{-1}$	de Vries et al. (1989)	(22)
$F_{glu-root}$	conversion factor of dry weight from glucose to root	0.857	kg ha^{-1} (kg ha^{-1})$^{-1}$	de Vries et al. (1989)	(23)
$F_{glu-starch}$	carbon fraction in the dry matter of starch	0.9	kg ha^{-1} (kg ha^{-1})$^{-1}$	Physical and chemical constant	(24)
K_N	vertical distribution of leaf nitrogen	0.11	-	Bonan et al. (2011)	(14)
r_{root}	rate of root depth increase	0.03	mm day^{-1}	Ordóñez et al. (2018) ; Nakano et al. (2021)	(11)
$Z_{rootmax}$	maximum root depth	1.7	m	de Vries et al. (1989)	(11)
T_b	base temperature for crop development	10	℃	de Vries et al. (1989)	(4)
T_{h}	highest temperature for crop development	34	℃	de Vries et al. (1989)	(4)
T_o	optimum temperature for crop development	27	℃	de Vries et al. (1989)	(4)

Table 2. Parameterized variables for soybean in MATCRO.

Variables	Value	Units	Description
a	-18.516	-	coefficient at relationship of rubisco activity and leaf nitrogen in Eq. (15)
b	114.33	-	coefficient at relationship of rubisco activity and leaf nitrogen in Eq. (15)
c	-73.336	-	constant at relationship of rubisco activity and leaf nitrogen in Eq. (15)
$DVS_{deadleaf1}$	0.6	-	1st DVS point where the dead leaf ratio pattern changes
$DVS_{deadleaf2}$	1	-	2nd DVS point where the dead leaf ratio pattern changes
DVS_f	0.4	-	developmental stage on initial flowering stage
DVS_{leaf1}	0.25	-	1st DVS point where the leaf partitioning pattern changes
DVS_{leaf2}	0.659	-	2nd DVS point where the leaf partitioning pattern changes
DVS_m	1	-	developmental stage at maturity time
DVS_{pod1}	0.48	-	1st DVS point where the pod partitioning pattern changes
DVS_{pod2}	0.72	-	2nd DVS point where the pod partitioning pattern changes
DVS_s	0.659	-	developmental stage to start seed filling stage
DVS_{SLN1}	0.4	-	1st DVS point where the specific leaf nitrogen changes along with DVS
DVS_{SLN2}	0.4	-	2nd DVS point where the specific leaf nitrogen changes along with DVS

Variables	Value	Units	Description
DVS_{SLN3}	0.659	-	3rd DVS point where the specific leaf nitrogen changes along with DVS
f_{starch}	0.18	-	fraction of glucose allocated to starch reserves
SR	0.68	-	seed-pod ratio (SR) accounting harvest index from storage organ
$N_{fert,high}$	300	$kgNha^{-1}$	nitrogen fertilizer value used in high nitrogen fertilizer in La Menza et al. (2023)
P_{leaf0}	0.38	-	glucose partitioning ratio of leaf toward shoot in the initial DVS point
P_{leaf1}	0.6	-	glucose partitioning ratio of leaf toward shoot in the 1st DVS point
P_{leaf2}	0	-	glucose partitioning ratio of leaf toward shoot in the 2nd DVS point
$P_{deadleaf1}$	0	s^{-1}	dead leaf ratio value in the 1st DVS point
$P_{deadleaf2}$	0.000001	s^{-1}	dead leaf ratio value the 2nd DVS point
SLN_{Y0}	0.75	gNm^{-2}	initial specific leaf nitrogen
SLN_{Y1}	2.25	gNm^{-2}	specific leaf nitrogen value in the 1st DVS point
SLN_{Y2}	1.7	gNm^{-2}	specific leaf nitrogen value in the 2nd DVS point
$SLN_{Y3,h}$	0.75	gNm^{-2}	specific leaf nitrogen value in the 3rd DVS point when using high nitrogen fertilizer
$SLN_{Y3,l}$	1.8	gNm^{-2}	specific leaf nitrogen value in the 3rd DVS point when using low nitrogen fertilizer
SLW_{max}	550	$kgha^{-1}$	maximum specific leaf weight
SLW_{min}	250	$kgha^{-1}$	minimum specific leaf weight
SLW_x	2.5	-	exponential slope of specific leaf weight to the developmental stage
$V_{ctopmax}$	103×10^{-6}	$mol(CO_2)m^{-2}s^{-1}$	maximum Rubisco capacity at the canopy top in Eq. (15)

3 Model calibration

The model's parameters were tuned using observed values for phenology and seasonality of biomass development. Once calibration was complete, the model continued to simulate crop growth, which encompasses phenological development, carbon assimilation, assimilate partitioning, and crop yield. We conducted calibrations to include various environmental conditions and soybean varieties documented in previous experimental studies as detailed in Sect. 3.1 and Table 3. The model calibration included parameterizing the dynamic biomass growth partitioning ratio for each organ (P_{organ}), leaf senescence, and specific leaf weight at each DVS. Other calibrations using the experimental dataset included the phenological stage, and the seed:pod ratio (SR). The crucial phenological stages (e.g., flowering and seed-filling) were calculated as the average value of the reported values in the experimental dataset. MATCRO applies this crop growth module following the method of the school od de Wit, comparing biomass growth with the observed values at various developmental stages. Shifts in partitioning and growth patterns were identified and used as reference points in the parameterization.

3.1 Description of site data used for calibration

The calibration process used experimental datasets reported in previous studies. The data were collected in field experiments across six different sites in four countries: Frederico Westphalen and Piracicaba (Brazil), Ya'an (China), Champaign (United States of America, US), Morioka and Tsukubamirai (Japan) (Table 3). The soybean cultivars grown at these experimental sites represented different maturity groups. A variety of management practices related to water management and nutrients were used in the field experiments. The farming practices differed among countries. The soybean plants were cultivated with a low

planting density in China and Japan, but typically at higher planting densities in the US and Brazil. Nitrogen fertilizers were applied at most sites, but the mineral nitrogen content in soil at sites in Brazil and the US was sufficient to support crop growth. Soybean crops were planted between May and June in the US, China, and Japan, but in October or November in Brazil.

Weather data were derived from the records at the meteorological station nearest to the experimental site. The climatic conditions at the respective sites were as follows: daily mean air temperature ranges during the growing season of 18–30 °C in Frederico Westphalen (Brazil), 19–31 °C in Piracicaba (Brazil), 17–27 °C in Tsukubamirai (Japan), 14–25 °C in Morioka (Japan), 18–26 °C in Ya'an (China), and 15–28 °C in Champaign (US). The seasonal precipitation (mm) was 1,669 mm in Frederico Westphalen (Brazil), 679 mm in Piracicaba (Brazil), 453 mm in Morioka (Japan), 865 mm in Tsukubamirai (Japan), 1,012 mm in Ya'an (China), and 787 mm in Champaign (US). The amount of solar radiation also differed among the experimental sites; China received the least solar radiation and Brazil received the most during the experimental period (Supplementary file Fig. S2). These data represent the diverse climatic conditions in soybean-producing countries. The field data used for calibration were collected across multiple crop seasons, specifically from 2002, 2003 to 2007 and from 2013 to 2016. These time periods were expected to capture the current climatic and environmental variability.

Table 3. Information about field experiments: Location, crop season, soybean variety and maturity group, water management, and nitrogen fertilizer, as well as the number of experiments for calibrating glucose partitioning ratio and evaluating soybean yield simulations.

Location	Crop season	Variety (RMG*)	Date of planting	Water management, Nitrogen fertilizer (g N m^{-2}), Plant Density (plant m^{-2})	Experiments (n)	Reference
Brazil (Frederico Westphalen)	2013	BRS284 (6)	1, 18 Oct; 8, 25 Nov, 12 Dec (2013)	Rainfed, 0, 26-28	5	Battisti et al. (2017)
Brazil (Piracicaba)	2013-2014	BRS284 (6)	18 Oct, 14 Nov (2013); 8 Jan (2014)	Irrigated and Rainfed, 0, 16-37	6	Battisti et al. (2017)
China (Ya'an)	2014	11 cultivars (5-8)	11 June (2014)	Irrigated, NA, 10	15	Wu et al. (2019)
	2014-2016	Texuan13 (7), Jiuyuehang (5), Nandou12 (6)	15 June (2014); 18 June (2015); 18 June (2016)		9	
United States (Champaign)	2002, 2004-2007	Pioneer93B15 (3)	1 June (2002); 28 May (2004); 25 May (2005, 2006); 22 May (2007)	Rainfed, 0, 25 – 53	8	Morgan et al. (2005); Ainsworth et al. (2007)
Japan (Tsukubamirai)	2013-2015	Enrei (2), Fukuyutaka (4), Ryuhou (2)	12 June, 31 July (2013); 17 June, 17 July (2014); 4, 30 June (2015)	Rainfed, 25-27, 9.5	16	Nakano et al. (2021)
Japan (Morioka)	2013-2016	Ryuhou (2)	13, 28 May (2013); 16, 30 May (2014); 5, 14, 25, 29 May (2015); 30 May, and 6, 27 June (2016)	Rainfed, 25-30, 9.5	10	Kumagai, (2018); Kumagai, (2021)

*relative maturity group

3.2 Biomass partitioning and specific leaf weight

The MATCRO-Soy model represents carbon assimilation by incorporating the carbon fraction in dry matter and that in glucose allocated to various plant organs. The glucose ratio for each organ was parameterized based on measurements of leaf weight, leaf senescence, stem weight, pod weight, and specific leaf weight across different developmental stages. To simulate glucose partitioning, we used Eqs. (25)–(29) to fit the segmented linear functions to the experimental dataset (Figure 2 and Figure 4) and the parameter values as shown in Table 2, as this value is used to obtain the average value of soybean partitioning behaviour.

The segmented linear functions for glucose partitioning were manually determined by visual inspections of the plot. This approach was chosen because of the challenges in applying nonlinear optimization. Breakpoints in the developmental stage were determined based on assumed growth characteristics, such as the decrease in leaf development after the seed-filling stage and the start of pod formation after flowering. We assumed an increasing trend of glucose allocation to leaf and shoot development during the early stage when data were unavailable, with subsequent segments aligned with observed data trends. The calibrated glucose partitioning ratio varied across the varieties and environmental conditions and was derived by converting biomass growth into glucose allocation as outlined in Eqs. (19)–(24).

The parameterization reflected the observed data, as well as the linear growth of leaves and pods during the developmental stages. It was used for seed:pod ratio and phenology parameterization. The dashed lines in Figure 2 and Fig. 3 indicate the estimated flowering and seed-filling stages, as determined by calculating the average time of phenological stages across all the experimental datasets. The independent dataset was used for evaluating the calibrated model at the point-scale level. After removing the calibration data, the simulated yield at the site scale showed a correlation coefficient of 0.68 and significant consistency (p value < 0.001) with observed data (Supplementary file Fig. S3). The simulated data were also consistent with observed data for aboveground biomass weight, pod weight, and leaf area index, with correlation coefficients of 0.60−0.90.

Assimilated carbon is subsequently allocated to other parts of the plant. Compared with varieties grown at other sites, the soybean varieties grown in Tsukubamirai (Japan) tended to have lower partitioning to the stem during the vegetative stage. The ratio of glucose to leaves in Sichuan (China) was unexpectedly high near maturity in 2016, resulting in a low level of partitioning to pods because of low temperature and drought conditions. The storage organ biomass increases during the reproductive stage to produce pods and seeds, whereas the shoot senesces at the end of the maturity period. Hence, yield is estimated using seed weight (as determined by storage organ weight) and the parameterized seed:pod ratio. In Champaign (US), pod partitioning tended to occur early during pod initiation in early maturing varieties. It is also observed in another study as the dry weight of pods before the seed filling stage was relatively high in early maturing varieties (Kawasaki et al., 2018). Early pod initiation occurred in the soybean variety ‘Ryuhou’ in Tsukubamirai in 2013 (Nakano et al., 2021). The dead leaf ratio parameter indicates the degree of leaf senescence after the seed-filling stage (Fig. 3), as calculated from the amount of leaf loss observed during the growing season.

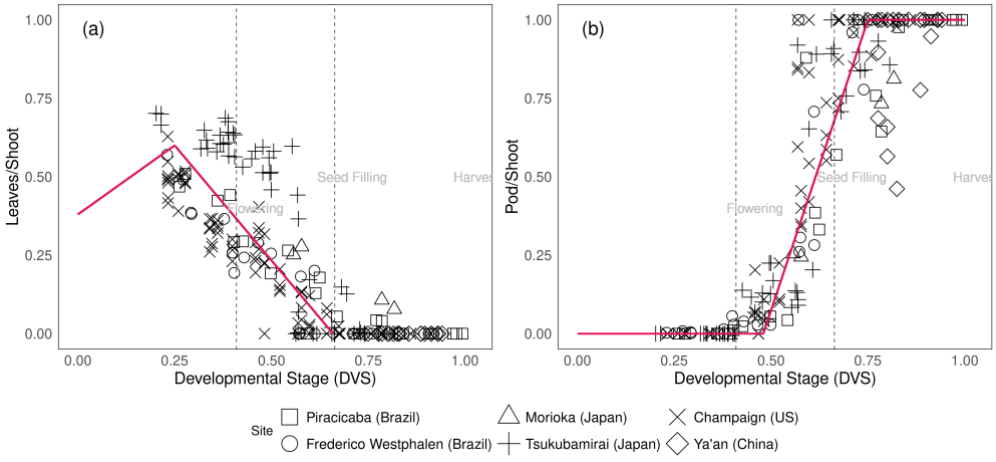

Figure 2. Glucose partitioning expressed as leaf:shoot ratio (a) and pod:shoot ratio (b) during soybean plant development [$DVS = 0$ (sowing) – 1 (maturity)] at different experimental sites (square: Piracicaba, circle: Frederico Westphalen, triangle: Morioka, plus: Tsukubamirai, cross:

Champaign, diamond: Ya'an. Red lines are segmented lines used for glucose partitioning in MATCRO-Soy. Dashed lines mark flowering, seed filling, and harvest times averaged across all experimental datasets.

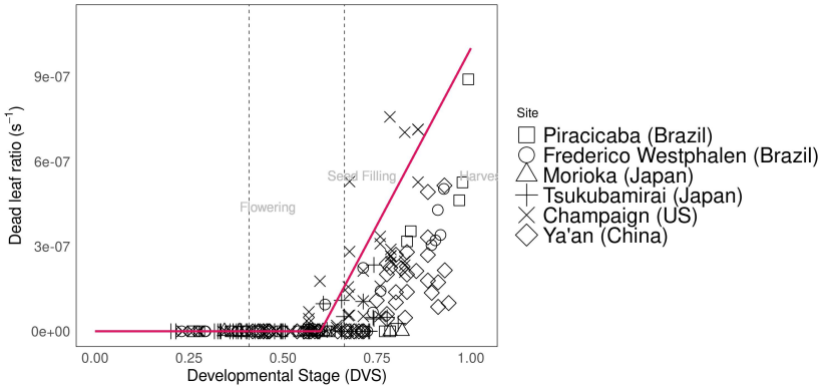

Figure 3. Dead leaf ratio (s−1) during soybean plant development (*DVS* = 0 – 1). Abbreviations and symbols are the same as in Fig. 2.

Specific leaf weight (*SLW*) is a significant parameter in crop growth parameterization and was calibrated to follow the measured data shown in Figure 4. We used the measured leaf weight and leaf area index data from the experimental datasets described in 2.4 and Eq. (30) to calculate the ratio of leaf weight to leaf area (SLW) during different phenological stages. These ratios change over time, and vary among growing seasons and cultivars (Thompson et al., 1996; Slattery et al., 2017). In the figure, SLW from Champaign (US) was excluded because of discrepancies in the timing of leaf area and leaf weight biomass measurements. While the SLW varied among the sites, we fitted the model of SLW assuming a saturating exponential function of developmental stage (red line in Figure 4). This pattern aligned well with the biological process, i.e., the SLW initially increases because of rapid biomass accumulation but saturates as the leaves mature.

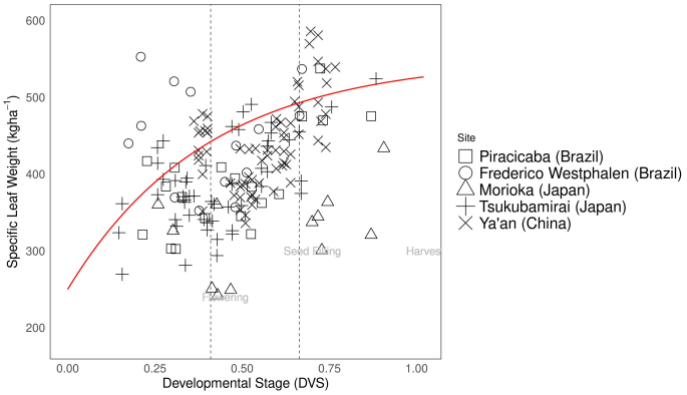

Figure 4. Specific leaf weight (kg ha^{-1}) during soybean plant development (*DVS* = 0 – 1). Abbreviations and symbols are the same as in Fig. 2.

4 Model evaluation setup

MATCRO was developed in FORTRAN and coupled with the global climate model's output, simulated at a spatial resolution of 0.5° × 0.5° and hourly–daily temporal resolution. The output of the model is gridded crop yield (kg ha^{-1}) as stored in netCDF file format in a global map with one harvest simulated per year. We evaluated the model's performance at global, country, and grid-cell levels for 34 years (1981–2014) at 0.5° spatial resolution with yearly harvested yield output. The accuracy of the simulated yield was assessed by comparison with reference global and country-level data from the Food and Agriculture Organization (FAOSTAT, 2023). The simulated grid-cell level yield was compared with Global Dataset of Historical Yield (GDHY) data, which are derived from statistical records, FAO data, and remote sensing data (Iizumi, 2019).

4.1 Simulation settings and data inputs

The parameters were set as shown in Table 4, covering the period of sowing years from 1980 to 2014, with various planting times across different regions. The model incorporated global daily climate data (86400 s) as input data. Although the simulation framework was that of the established MATCRO-Rice v.1 (Masutomi et al. 2016b), several modifications were made to enhance its applicability at a global scale. Notably, the temporal resolution was adjusted from half-hourly (1800 s) to hourly (3600 s), allowing the model to maintain consistency in capturing critical processes such as diurnal variations in photosynthesis and transpiration, while optimizing computational efficiency. These adjustments ensured that the model was suitable for large-scale simulations while preserving essential physiological processes.

The model simulates soybean yield using input data as described in Table 5. It uses the following global input data: crop calendar from the Global Gridded Crop Model Intercomparison (GGCMI), which separates rainfed and irrigated systems; atmospheric CO_2 and climate data from the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP), which provides bias-adjusted climate input data for historical data (GSWP3-W5E5 v2.0); soil classifications from the Harmonized World Soil Database (HWSD v1.2); and nitrogen fertilization for C_3 fixing crops of the ISIMIP, which is derived from the land use dataset (Hurtt et al., 2020). ISIMIP bias-adjusted data are used to maintain uniformity in the climate impact data across sectors and scales in the framework. This dataset, which is provided by ISIMIP, has a spatial resolution of 0.5°. To determine the growing degree days for maturity, we considered the phenological maturity time from the GGCMI crop calendar for harvest time and global ISIMIP climate data over 10 years (2000–2010) to capture the shifts in variability across the current evaluation years.

Table 4. Parameter settings for simulation.

Variable	Value	Unit	Description
$Year_{sow}$	varied	Year	year of sowing day
DOY_{sow}	varied	Day of Year (DOY)	day of year of sowing day
RES_{time}	3600	s	time resolution for simulation
$RES_{climate}$	86400	s	time resolution for climate forcing data
$RES_{we/ns}$	0.5	degree	spatial resolution north to south or west to east
Soil layer	5.0	-	number of simulated soil layer to calculate soil water content
WSL	1.0	-	soil water content at emergence
W_{leaf0}	1.0	kg ha^{-1}	dry weight of leaf at emergence
W_{stem0}	1.0	kg ha^{-1}	dry weight of stem at emergence
W_{root0}	1.0	kg ha^{-1}	dry weight of root at emergence
W_{glu0}	0.5	kg ha^{-1}	dry weight of glucose reserve at emergence
Z_a	3.0	m	reference height at which wind speed is observed
Z_{max}	4.0	m	depth of soil layer
Zt	0.05	m	depth of topsoil layer
Z_b	2.0	m	depth from the soil surface to the upper bound of the most bottom layer of soil

395

Table 5. Data input for MATCRO simulation.

Variable	Unit	Data source	Spatial resolution
Daily time-step			
Precipitation	kg $m^{-2}s^{-1}$	GSWP3–W5E5 (Kim, 2017; Cuchi et al., 2020; Lange, 2019; Lange et al., 2021)	0.5° × 0.5°
Near-surface specific humidity	kg kg^{-1}		0.5° × 0.5°
Maximum, minimum, and mean temperature	Kelvin		0.5° × 0.5°
Surface downwelling shortwave radiation	W m^{-2}		0.5° × 0.5°
Near-surface wind speed	m s^{-1}		0.5° × 0.5°
Surface air pressure	Pa		0.5° × 0.5°
Yearly time-step			
Atmospheric CO2 concentration	ppm	ISIMIP (Büchner and Reyer, 2022)	-
Nitrogen fertilizer	kg ha^{-1}	ISIMIP (Volkholz and Ostberg, 2022)	0.5° × 0.5°
Constants			
Latitude and longitude	°	-	-
Agricultural management	Irrigated or rainfed	MIRCA2000 (Portmann et al., 2010)	0.5° × 0.5°
Sowing time, Harvest time	DOY	GGCMI (Jägermeyr et al., 2021)	0.5° × 0.5°
Growing degree days for harvest time	°C days	Parameterized in this study	0.5° × 0.5°
Soil type	-	HWSD (Volkholz and Müller, 2020)	0.5° × 0.5°

4.2 Global yield evaluation methods

In this study, we assessed the statistical relationship between simulated yields and observed or reference data using the common metrics of Pearson's correlation coefficient (corr) as calculated using Eq. (33) with significance level (*p*-value), agreement between the simulated and observed results using root mean square error (RMSE) as calculated using Eq. (34), and relative bias, as calculated using Eq. (35), for time-series yield data as follows:

$$corr = \frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2(Y_i-\bar{Y})^2}} \quad (33)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - Y_i)^2} \quad (34)$$

$$relative\ bias = \frac{1}{n}\sum_{i=1}^{n}|X_i - Y_i| \times \frac{1}{\bar{Y}} \quad (35)$$

where X_i and Y_i indicated simulated and observed values for each measurement; $\bar{X}$ and $\bar{Y}$ denote the mean of simulated and observed values for the harvested year annually; and i and n are the i-th data point and total number of data, respectively. We used n = 34 years for global-scale data, while the output after calibration was evaluated at the point-scale using n = 14–122 of the available experimental datasets.

Detrended yield represents the time-series yield data for both simulated and observed values after removing the linear trend by subtracting the slope and intercept of the fitted linear regression (long-term yield trend). This approach enables the

separation of short-term yield fluctuations from systemic long-term shifts. To provide a clear interpretation of the model's evaluation errors, yield fluctuations were evaluated separately for the long-term and detrended data using mean squared deviation (*MSD*) and its components (Gauch et al., 2003; Kobayashi and Salam, 2000), as outlined in Eq. (36):

$$MSD_y = SB_y + SDSD_y + LCS_y \quad (36)$$

where MSD_y is the square of RMSE for each long-term yield trend or detrended yield. Its components include mean squared bias (SB_y), difference in the magnitude of fluctuation, namely the squared difference between standard deviations ($SDSD_y$), and the lack of positive correlation weighted by the standard deviations (LCS_y) as proposed by Kobayashi and Salam (2000). These terms were calculated using Eqs. (37)–(41), as follows:

$$SB_y = (\bar{X} - \bar{Y})^2 \quad (37)$$

$$SDSD_y = (SD_X - SD_Y)^2 \quad (38)$$

$$SD_X = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2} \quad (39)$$

$$SD_Y = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2} \quad (40)$$

$$LCS_y = SD_X SD_Y(1 - corr) \quad (41)$$

Higher SB_y, $SDSD_y$, and LCS_y indicate that model failed to simulate the mean of the measured yield, magnitude of fluctuation around the mean yield, and pattern of fluctuation in yield across n measurements, respectively. SD_X and SD_Y denote the standard deviation of simulated (X) and observed values (Y), respectively, and LCS_y depends on the correlation coefficient (corr).

5 Evaluation of model performance

We calculated soybean yield with a global-scale map based on the gridded data of irrigated and rainfed area from the MIRCA2000 dataset, which represents global agricultural land use around the year 2000 (Portmann et al., 2010), to get the actual yield value. We evaluated yield during the period of 1981–2014 because the MIRCA dataset was available within that period. The simulated yield at the global scale, and at the country scale for regional comparison, was determined by aggregating grid cell data to compute the mean soybean harvested area within each country grid as described below in Eq. (42):

$$Yield_{region} = \frac{\sum_{i=1}^{n}[(Yield_{rf})_i(Area_{rf})_i + (Yield_{ir})_i(Area_{ir})_i]}{\sum_{i=1}^{n}[(Area_{rf})_i + (Area_{ir})_i]} \quad (42)$$

where $Yield_{region}$ is the aggregated yield in a given region (country or global-scale) in kg ha^{-1} from the grid cell number (i) ranging from 1 to n (total number of grid cells in the region); $Yield_{rf}$ and $Yield_{ir}$ are estimated yield under rainfed and irrigated conditions,and $Yield_{ir}$ respectively; and $Area_{rf}$ and $Area_{ir}$ are the soybean rainfed and irrigated area (ha),and $Area_{ir}$ respectively.

5.1 Model output yield as evaluated at the global and national scales

Figure 5a shows a time-series comparison from 1981 to 2014 between the global mean yields reported by FAOSTAT and those simulated by MATCRO-Soy. The results show that the model captured the upwards trend in yields over the period with a shallower slope compared with that of the reported yield data. The correlation coefficient was 0.81, and was significant ($p < 0.01$); and the errors were RMSE of 298 kg ha^{-1} and relative bias of 0.12. Notably, the simulated linear increase contributed to the higher correlation coefficient for the yield trends.

Figure 5b compares the detrended global mean yield observed by FAOSTAT and the simulated value by MATCRO-Soy after removing the long-term linear trend across the study period. The detrended yield is the value after the long-term trend is subtracted from the original yield data. It isolates the variability primarily driven by climate fluctuations to evaluate interannual variability independent of long-term trends. However, it also removes longer-term signals (e.g., effects of technological improvements or increasing CO_2 concentrations). The correlation coefficient for the detrended yield data decreased to 0.446 ($p < 0.01$). The model reproduced the interannual variations well with an RMSE of 137 kg ha^{-1}. Specifically, according to observed data, there were significant yield reductions in the years 1983, 1988, 2009, and 2012. Among these, the model successfully reproduced the yield reductions in three years (1983, 1988, and 2012), excluding 2009. Severe droughts occurred in those years, and the model's ability to capture these events is noteworthy.

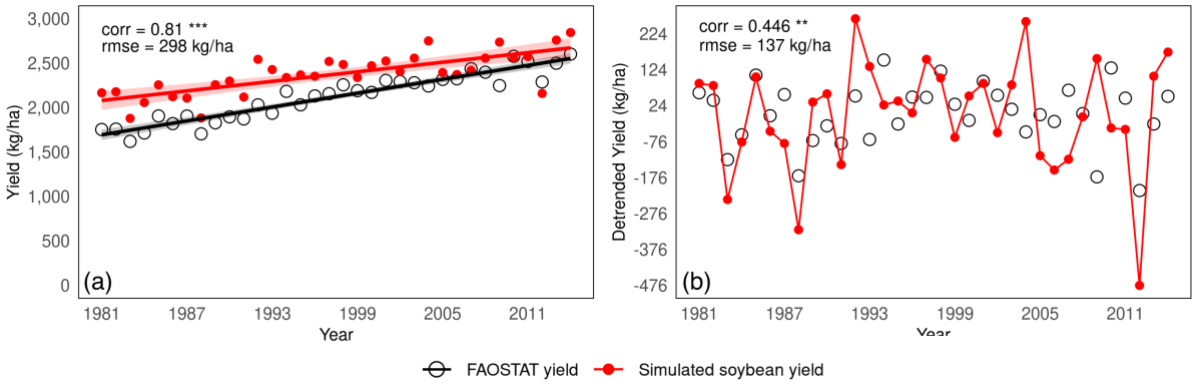

Figure 5. Time-series comparison between simulated yields by MATCRO-Soy and FAOSTAT reported yield data: (a) Global yield and long-term trend during 1981–2014, and (b) Detrended yield during 1981–2014. Correlation for detrended yield was calculated by subtracting linear trend. Symbols ***, **, and * denote $p < 0.001$, 0.01, and 0.05, respectively.

We evaluated the model's performance for 10 major soybean-producing countries; Argentina, Brazil, China, India, Paraguay, the US, Italy, Russia, Bolivia, and Canada, which together account for 96 % of global soybean production (based on total average production from 2012 to 2021 reported in FAOSTAT). Figure 6 compares the simulated average yields per country and the reported average yields per country as reported in FAOSTAT for 1981–2014 with the ellipsoids indicating the distribution of the simulated yield values within the 90 % confidence range. The model reproduced the national average yield levels well for the top 10 producing countries, as indicated by a correlation coefficient of 0.519 ($p < 0.001$) and an RMSE of 1,085 kg ha^{-1}. The correlation coefficients were significant for six countries; Argentina, Brazil, India, Italy, Paraguay, and the US (see Supplementary file Fig. S4 for further evaluation of these six countries). Focusing on the US, Brazil, and Argentina, which together account for 69 % of global soybean production, the model's accuracy showed a correlation coefficient of 0.645 ($p < 0.001$) and an RMSE of 916 kg ha^{-1}, although soybean production in Brazil was underestimated. When all 10 countries

were considered, the correlation coefficient decreased to 0.291, although it remained statistically significant. These results
470 demonstrate that the model performs reasonably well in capturing yield variations in major producing countries and achieves particularly lower bias in some countries (e.g., the US, Italy, and Canada).

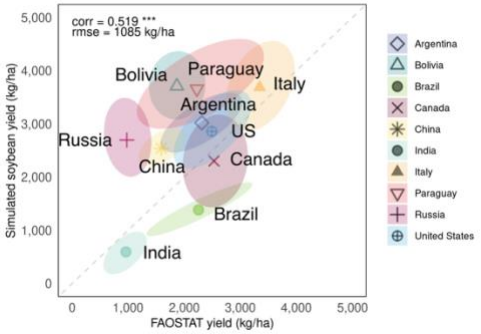

Figure 6. Comparison between average yields simulated by MATCRO-Soy and average yields reported by FAOSTAT for the 10 major soybean-producing countries during 1981–2014. Ellipsoid shows 90 % confidence range of annual yield.

475 A time series comparison of average yields for each of the 10 major soybean-producing countries is shown in Figure 7. An evaluation of the long-term trends (Figure 7a) revealed that MATCRO-Soy effectively captured the trends of increasing soybean production. The modelled and observed trends showed the strongest agreement in Brazil, followed by Argentina (0.62) and the US (0.64). The detrended yields revealed interannual variability (Figure 7b). For these data, the modelled and observed data had the highest correlation coefficient in Paraguay (0.61), followed by the US (0.57) and Brazil (0.49), and the
480 lowest correlation coefficients in China (0.18) and Bolivia (−0.32). These findings suggest that the model tends to perform with greater accuracy for countries with higher production levels, even in time series comparisons at the national level.

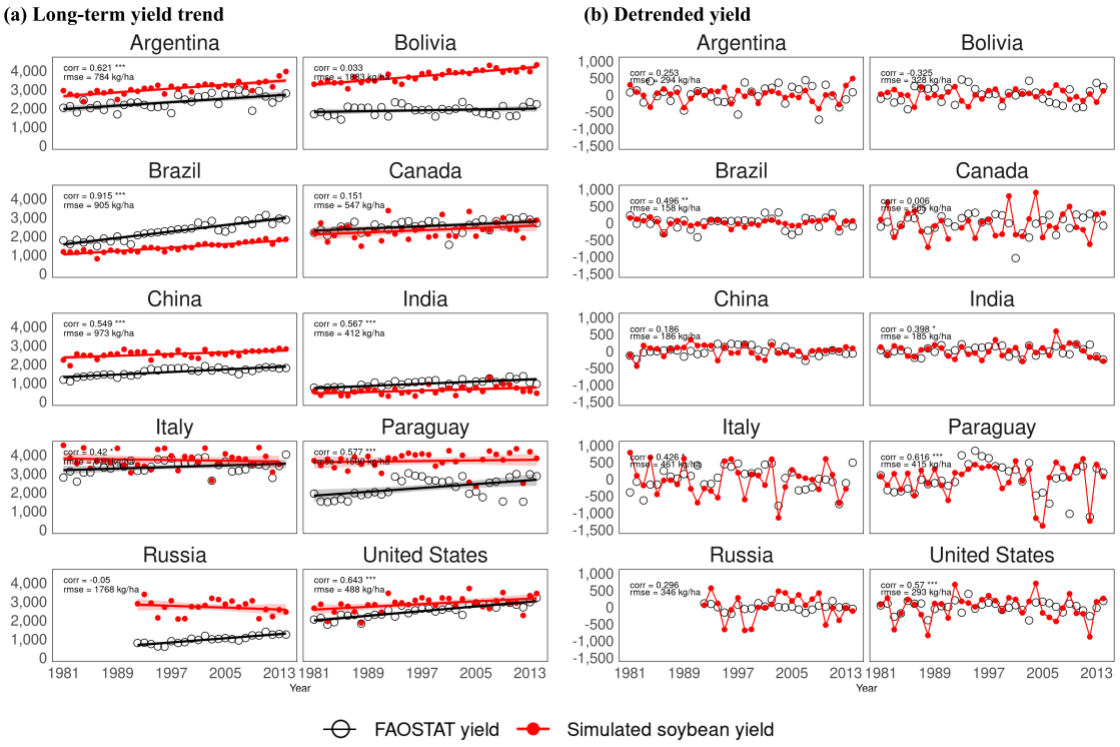

Figure 7. Time-series comparison between yields simulated by MATCRO-Soy (red circle) and yields reported by FAOSTAT (open circle) in 10 top soybean-producing countries during 1981–2014: (a) Long term yield trend in kg ha^{-1} (solid line), (b) Detrended yield in kg ha^{-1} after subtracting linear trend. Correlation coefficient and RMSE are shown in each panel. Symbols ***, **, and * denote $p < 0.001$, 0.01, and 0.05, respectively. Shading near solid line is standard error with confidence interval of 95 %.

5.2 Temporal trends and variability

The model's performance was further assessed with the *MSD* components for yield, separated into yield, long-term yield trend, and detrended yield for both the global (Supplementary file Table S1) and country scales (Supplementary files Tables S2, S3, and S4). We separated the *MSD* into *SB, SDSD*, and *LCS*, which reflect errors in mean yield, magnitude of yield variability, and the pattern of year-to-year fluctuations, respectively. The greatest contributor to error at the global scale was *SB*, contributing approximately 71 % and 77 % of total *MSD* for yield and detrended yield, respectively (Supplementary file Table S1).

Figure 8 shows *MSD* components in the top six soybean-producing countries. In most countries, *SB* was the primary source of error. The highest *MSD* was in Paraguay, and was largely driven by *SB*, with a notable contribution from *LCS*. This indicates that the model simulated variability well but poorly captured the mean yield. The low *MSD* in the US was also driven by *SB*, but *LCS* also contributed to year-to-year variability. Meanwhile, *LCS* was the greatest contributor to yield error in Canada and Italy (Supplementary file Table S2) because of pronounced discrepancies in the simulated interannual variability. *SDSD* contributed to error only in Brazil, where the model underestimated the mean yield and the deviation. These results highlight that the mean yield bias is main source of error at the global and country scales, while *LCS* and *SDSD* contribute notably in specific regions where the model fails to capture the variability or the temporal pattern.

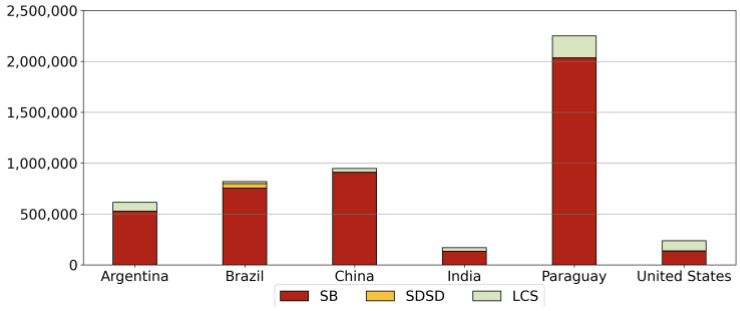

Figure 8. Mean squared deviation (*MSD*) components of squared bias (*SB*), sum of difference in standard deviation (*SDSD*), and lack of positive correlation (*LCS*) for yield error in top six soybean-producing countries.

5.3 Model performance at the grid-cell level

We evaluated MATCRO-Soy at the grid-cell level by comparing simulated yields with observed ones from the GDHY dataset reported by Iizumi (2019). Figure 9a and b show the simulated and observed yields averaged over 34 years, and Figure 9c shows relative bias between them. Figure 10 shows the interannual correlation between simulated and observed yields for 34 years. The simulated yield was calculated for soybean-growing areas from the MIRCA2000 dataset, which offers broad spatial coverage where yield data for certain regions, including Canada, Russia, Australia, and many European and Asian countries, are missing in the GDHY dataset (Iizumi and Sakai, 2020). The density plot of the simulated yield showed more variability than did the GDHY data in Figure 9. However, both datasets exhibited a density peak of approximately 2,000 to 3,000 kg ha^{-1} and the simulated yield mostly overestimated the higher yield value. Figure 9 a, b, and c also show the distribution of simulated and observed yields.

The relative bias map (Figure 9c) highlights that overestimation was prominent in parts of South America (particularly Argentina), Russia, and China. In contrast, the model tended to underestimate yields in South Africa, India, and Brazil. Most of the grid cells in Brazil showed low yields, likely due to shorter growing periods in the input data compared with those in the field experimental data. These results aligned with the trends observed at the national scale, which were influenced by the aggregation process. During aggregation, the national-scale results represent the average performance across all grid cells, weighted by the number of grids within each region. Most grids had a relative bias of −0.2 to 0.2, accounting for 37 % of the total grid area. For areas shown in grey, the correlation was statistically insignificant. The density plot of simulated yield showed more variability compared with the GDHY data. However, both datasets exhibited a density peak at approximately 3,000 kg ha^{-1}, and the simulated yield mostly overestimated the observed yield. Correlation coefficients were calculated for each grid cell between the simulated yield and the GDHY dataset after removing the moving-average to reveal interannual variation (Figure 10). The correlation was significant in 66 % of the grid cells ($p < 0.05$).

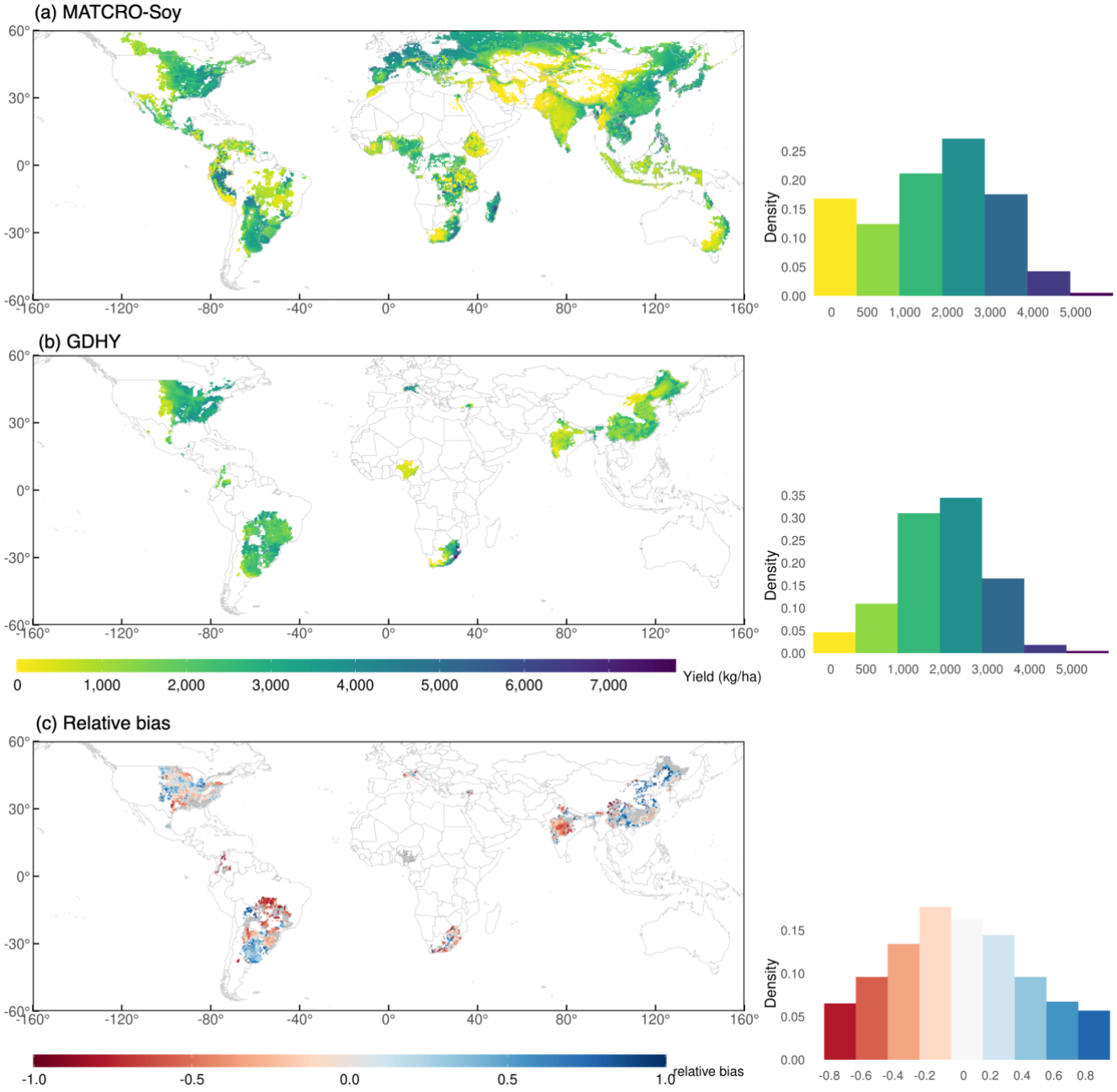

Figure 9. Global map of 34-year averaged (1981–2014) yield of GDHY dataset (a), simulated by MATCRO-Soy (b), and relative bias (c) with each density plot distribution. In (c), areas in grey are those where the correlation between simulated and observed yields was non-significant ($p > 0.05$).

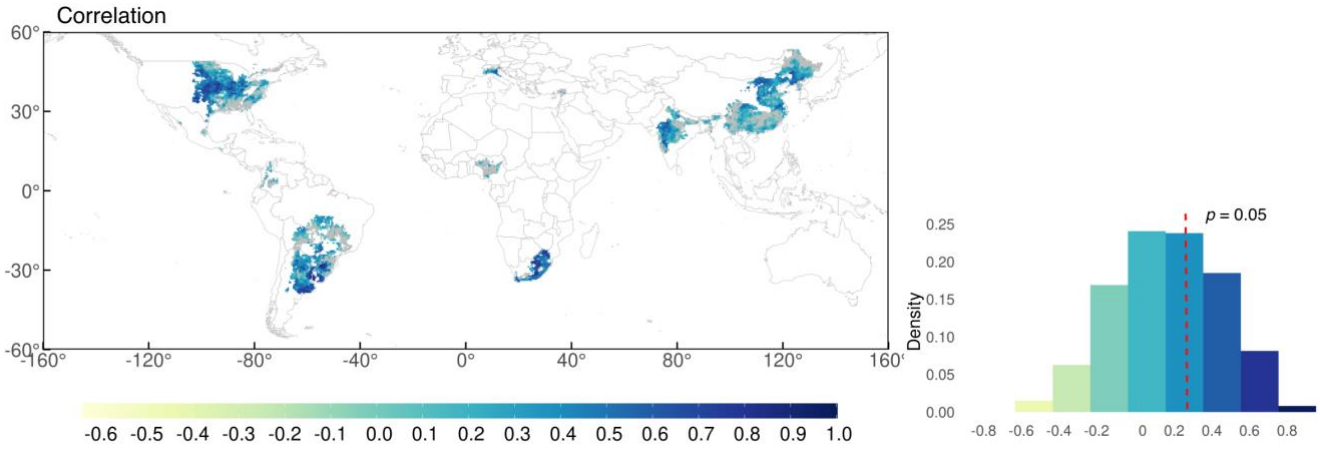

Figure 10. Time-series correlation between simulated and observed yield in 1981–2014 after removing trends from 5-year moving average (c). Grey colour depicts regions with non-significant correlations ($p > 0.05$) in the map. Red dashed line shows the border of $p = 0.05$ for the number of years ($n = 34$) in the density distribution plot.

Figure 11 presents the relative RMSE (RMSE value compared with the observation value) between the simulated yields and GDHY datasets for detrended yield at the grid-scale. The RMSE values were relatively higher in some parts of Africa (particularly in Nigeria), the US, India, and China, and relatively lower in Brazil and Argentina. India and the US showed low RMSEs at the national-level, but some grid cells within both countries had higher relative RMSEs at the grid-cell level. Detailed information on the spatial variation in error and its components is provided in Supplementary File Fig. S6 for the long-term yield trend and Supplementary File Fig. S7 for the detrended yield.

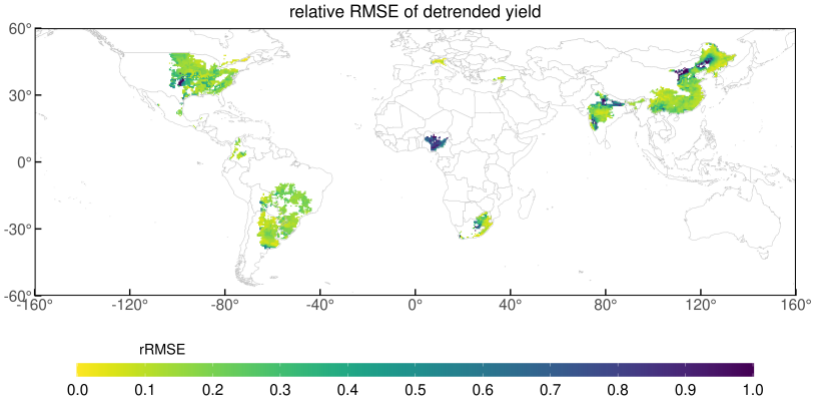

Figure 11. Relative RMSE calculation between simulated and observed yield for detrended yield at the grid-cell level.

5.4 Model performance at the leaf-level

We simulated the leaf-level variation in V_{cmax} at the site-scale for Champaign, US (the country with the highest soybean production), for the 2002 growing season. The global parameterization of MATCRO-Soy was used for this simulation (Figure

12). The leaf-level simulated V_{cmax} values aligned closely with the measured data reported by Bernacchi et al. (2005) during the vegetative stage, but showed some deviations during the flowering to seed-filling stages (dotted line in Figure 12). This alignment highlighted the ability of the model to represent essential photosynthetic processes influenced by leaf nitrogen content.

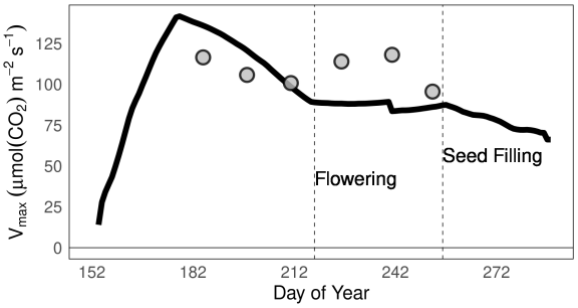

Figure 12. Maximum carboxylation capacity of Rubisco ($\mu mol(CO_2)m^{-2}s^{-1}$) during the growing period in Champaign (US) in 2002 as simulated using MATCRO-Soy (black line) and as measured (grey dots) by Bernacchi et al. (2005).

6 Discussion

6.1 Validation of MATCRO-Soy

In prior studies, soybean yield predictions often faced challenges in capturing crop responses to climatic variables. Our results show that the MATCRO-Soy model effectively captures the linear trend in soybean yields, with higher accuracy for long-term trends (corr = 0.812) than for detrended yields (corr = 0.446) (Figure 5). This result for the global detrended yield improves upon that of the benchmark study of Müller et al. (2017), indicating that there is less variation in process-based models based on their statistical correlations. Another crop model, PRYSBI2, reached a significant correlation of 0.57 ($p < 0.050$) based on long-term trends. However, model accuracy is enhanced when site-specific parameters are used. This has been demonstrated in regional-scale evaluations in previous studies, which were used for parameterization in this global simulation (Battisti et al., 2017; Kumagai, 2018, 2021; Morgan et al., 2005; Nakano et al., 2021; Wu et al., 2019). Those studies showed that integrating factors such as cultivar differences, ensembles of multiple crop models, nitrogen content, and more accurate measurement methods allow for a more reliable representation of local growing conditions and climate variability.

We examined the model's performance in predicting soybean yields for the 10 largest soybean-producing countries. As shown in Figure 6, the RMSE was 1,085 kg ha^{-1} (average yield over 34 years). This value is similar to that reported in another study using the LPJ-GUESS model incorporating a biological nitrogen fixation module (Ma et al., 2022), where the RMSE was approximately 800 kg ha^{-1} (average yield over 10 years). Evaluation of MATCRO-Soy's performance at the grid-cell level, as shown in Figure 9, revealed that the correlation between simulated and observed yields was significant ($p < 0.05$) in 66 % of the grid cells, with the value in most grid cells within the range of 0.2–0.6. These findings align with those of the benchmark study of Müller et al. (2017), who found that time-series correlations in GGCM-simulated soybean yields ranged from 0.25 to 0.65 because of various discrepancies in the data. These correlations were calculated using detrended values, which is a useful strategy for evaluating interannual variability and the model's sensitivity to climate fluctuations. However, detrending removes important long-term signals related to genetic improvements, cultivar and management changes, and/or increased CO_2 levels.

Analyses of the correlations between yield and detrended yield (Figure 5 and Figure 6) indicate that the model performed better (i.e., a higher correlation coefficient) when predicting long-term yield trends. MATCRO-Soy was able to capture the trend of increased atmospheric CO_2 and nitrogen fertilizer inputs, despite the interannual variability in climate conditions. Analyses of *MSD* and its components revealed that the lack of positive correlation was the major contributor to error in Canada and Italy among the 10 top soybean-producing countries (Supplementary file Table S2). The *SB* values were small for both Canada and Italy, suggesting that MATCRO-Soy accurately represents the average productivity despite its inability to capture the variability or amplitude of the yield trend over time within those regions. Factors such as changes in sowing date, land use, pest management, cultivar maturity group, and planting density may contribute to discrepancies in soybean yield under climate change (Battisti et al., 2018a; Marin et al., 2022). Hence, there is a need for improved parameterization to better represent the dynamics of yield variability in countries such as Canada and Italy.

The high yields in Argentina and Paraguay reflect the consistency of favourable growing conditions in those countries (Figure 7a), particularly the alignment of daily temperatures and seasonal precipitation with critical growth stages. This result suggests that these regions are less susceptible to interannual variability, as well as being located in areas that receive more radiation for photosynthesis. The comparison of simulated and observed yields at the grid-cell level (Figure 10) revealed weak correlations with no statistical significance in high-latitude countries with a low number of grid cells (e.g., Canada and Russia). Models that do not include daylength have a higher level of uncertainty (Battisti et al., 2018b). Moreover, the low simulated yield in India, which has a hot climate characterized by high mean daily temperatures of 27–28 °C (Supplementary file Fig. S5) and low soil moisture during the growing season, highlights the capacity of the model to capture regional climatic challenges that impact productivity. These climatic challenges likely exacerbate heat stress during critical phenological stages, such as flowering and pod development, leading to reduced yields (Sinclair, 1986; Egli and Bruening, 2004). The contrasting regions of high and low soybean yields underscore the ability of the model to capture the complex interplay between climate and crop yields across diverse agroecological zones.

6.2 Model strength and application

MATCRO-Soy v.1 was developed as a process-based eco-physiological model that uses the Farquhar equation to simulate leaf-level photosynthesis. The Farquhar equation is a widely recognized framework in plant physiology that simulates the biochemical mechanisms of photosynthesis by describing the relationships among light intensity, CO_2 assimilation, and Rubisco activity (Farquhar et al., 1980; Scafaro et al., 2023). Through the integration of this equation into a gridded global crop model, MATCRO-Soy enhances the simulation of soybean growth and productivity under environmental changes in atmospheric CO_2 levels, temperature, and water availability. These factors are important for predicting and understanding how climate changes affect productivity. The calibration of MATCRO-Soy successfully represented the response of soybean plant growth to a wide range of climatic conditions, resulting in reliable global yield simulations using a single parameterization. While simplification may introduce errors, global tuning effectively minimizes these discrepancies in specific regions. This conclusion was also observed in Smith et al. (2014).

Improving photosynthetic efficiency is a key goal for crop improvement, particularly through enhancing stomatal conductance and modifying Rubisco, the enzyme responsible for carbon fixation (Xu et al., 2022). We used V_{cmax} as a photosynthetic parameter because it quantifies the activity of Rubisco, which catalyzes the conversion of CO_2 into organic compounds. The peak Rubisco activity during the reproductive stage corresponds with trends in *SLN*, and is implicitly affected by additional nitrogen fertilizer (La Menza et al., 2023). It is also important to consider nitrogen fixation, because it is reduced under adverse environmental conditions such as flooding, water deficit, and low temperatures (Santachiara et al., 2019).

Prior to the global scale evaluation, the yield, *LAI,* aboveground biomass, and pod biomass simulated by MATCRO-Soy were further compared at the point-scale level with experimental datasets, with distinct datasets used for each step of calibration and evaluation (Supplementary file Fig. S3). While point-scale simulations employed the unified global parameters, the results

demonstrated reasonable agreement with $p < 0.01$ and bias of 30 %–63 % for harvested yield, seasonal *LAI*, aboveground biomass, and pod biomass. The highest bias was observed for seasonal *LAI*, which aligns with the underestimation of V_{cmax} during critical growth stages. Thus, MATCRO-Soy can reproduce photosynthesis parameters comparable to those of observed data at the site scale, although it overestimates these parameters at the reproductive stage (Figure 12).

MATCRO-Soy uses high-quality data related to climatic factors, soil, and nitrogen fertilization to capture biophysical processes involved in soybean growth and yield formation. Its flexibility in terms of spatial resolution allows it to be applied across various scales, from grid-level to country to global. Moreover, MATCRO is easily coupled with climate and atmospheric CO_2 models to increase the accuracy of yield predictions through high-quality data inputs. This adaptability also enables integration with other land models, making it a valuable tool in both ecological and agricultural research. MATCRO-Soy can be continuously refined with new data and plant physiological knowledge, ensuring that it remains robust and adaptable. This adaptability makes it a valuable tool for researchers and policy-makers working towards sustainable agriculture and global food security.

The strength of MATCRO-Soy lies in its ability to simulate key physiological processes of soybean growth (e.g., photosynthesis, phenology, and biomass partitioning) under varying climatic conditions. Its process-based structure allows for sensitivity analysis for evaluation of further environmental impacts, such as effects of elevated CO_2 and temperature stress. MATCRO-Soy reasonably captures the temporal dynamics of yield formation. Fluctuations in yield are influenced not only by climatic conditions, but also by advances in technology, evolving agricultural practices, and modifications to crop management approaches. Although these impacts are outside the scope of model, their inclusion can further improve accuracy at the local scale. For example, including pest and crop interactions may enhance the model's capability to reflect local yield responses to climate-driven pest dynamics (Chen and Mccarl, 2001). The integration of crop models with remote sensing data will enhance their capability for monitoring and predicting crop productivity at finer spatial scales (Basso et al., 2001). However, it is important to acknowledge the limitations of the MATCRO-Soy model, particularly its ability to predict yield variations under extreme or rapidly changing climatic conditions. Continuous updates of the experimental dataset are necessary to maintain its relevance and accuracy in predicting future soybean yields.

6.3 Model challenges and future directions

In the evaluation process, we observed considerable interannual variability and spatial variability. In Brazil, there were many grid cells with a low, non-significant correlation between simulated and observed soybean yields (Fig. 9), but the correlation at the national-scale level was high (Figure 7). This means that local climatic factors affect soybean yield in Brazil. However, MATCRO-Soy recognizes broader regional trends, fulfilling its aim to represent yield behaviour (Figure 11). These findings highlight that the number of grid cells significantly influences the model's performance, with regions containing fewer grids being more sensitive to localized factors and spatial heterogeneity during aggregation. This emphasizes the importance of considering spatial resolution and representation when evaluating model performance.

Uncertainty in MATCRO-Soy is reflected by the challenges in evaluating the model at a global scale, particularly due to its assumption of globally homogeneous crop cultivars and the upscaling processes due to limited observed data. This means it is unrealistic to reproduce variability at the regional scale with high accuracy (Müller et al., 2017; Zaehle and Friend, 2010). This uncertainty is notably pronounced in the global aggregation of yield simulations at the grid-cell scale. Global aggregation can escalate substantially for specific combinations of aggregation units, crop model limitations, and years (Porwollik et al., 2017). Future assessments of models and projections of crop yields will require careful consideration of the significant contrast between different aggregation approaches used for individual countries or regions. To address this, we used harmonized ISIMIP data to minimize methodological bias, and developed the model with sufficient flexibility to reduce uncertainty (Yin, 2013).

Comparison of soybean yields simulated using bias-corrected climate data with FAO data revealed a large underestimation in 2002 and an overestimation in 2009 (Figure 5). One possibility for these discrepancies in interannual variability is the influence of not accounted for extreme climatic events. Climatic events indicated by the Oceanic Niño index, a three-month running mean of SST anomalies in the Niño 3.4 region, show that La Niña was present at the end of 2002 and that El Niño occurred at the end of 2009 (NOAA, 2024). Some regions within major soybean-producing countries are significantly affected by El Niño events, further influencing yield variability (Anderson et al., 2017; Iizumi et al., 2014). Another reason why MATCRO-Soy tends to overestimate long-term yield trend is that its carbon assimilation module is sensitive to changes in the atmospheric CO_2 concentration.

While nitrogen fixation and uptake are implicitly constrained by the *SLN* parameter, the carbon cost economic approach explicitly represents the respiratory cost due to different nitrogen uptake pathways (Fisher et al., 2010). MATCRO-Soy simplifies the nitrogen fixation mechanism, and this may have contributed to yield overestimation in countries with low nitrogen inputs (e.g., Bolivia and Russia). However, the model still showed a relatively small bias in countries with high nitrogen fertilizer application (e.g., China), as well as in countries with low nitrogen fertilizer input (e.g., the US). This highlights an opportunity for future model development to incorporate a variable for the respiratory costs of biological nitrogen fixation. There are limited empirical data across cultivars, environments, and management systems, and this poses a challenge for yield predictions at the global scale. Further experiments on the respiratory costs of nitrogen fixation would improve our understanding of the physiological mechanisms of soybean plants under nitrogen-limited conditions.

Simulated yield increases throughout the year driven by the positive effects of increased atmospheric CO_2, a phenomenon known as the CO_2 fertilization effect, were reported in studies by Long et al. (2005) and Sakurai et al. (2014). The CO_2 fertilization response may become a more prominent source of overestimation in future projections if the model overestimates the crop response to elevated CO_2. Compared with simulations using statistical radiation use efficiency (Ai and Hanasaki, 2023), process-based models have this tendency because of the greater effect of CO_2 on photosynthesis. Therefore, further investigation is needed to fine-tune the CO_2 sensitivity of MATCRO-Soy and other process-based models, because photosynthesis is known to be downregulated under elevated CO_2 (Ainsworth et al., 2002; Zheng et al., 2019). This is especially important for adaptation studies, as reliable yield projections are critical for designing effective adaptation strategies under future climate scenarios.

Analyses of MATCRO-Soy simulation errors showed that the *MSD* component *SB* was the dominant contributor to errors in yield prediction at the global and country scales. This indicates that the bias was in the over- or underestimation of average yield, rather than in yield variability or the year-to-year yield pattern (Figure 8). These results highlight the model's uncertainty in simulating mean yield in major soybean-producing countries with large cultivation areas. The model overestimated the long-term yield trend in some countries. Inaccurate representation of CO_2 fertilization effect may have contributed to the mean yield bias. Other factors that may contribute to this bias are the simplified assumption of no respiratory costs for symbiotic nitrogen fixation and insufficient representation of water stress responses. The accuracy of data inputs may partly reflect the inherent gap between field experiment data and national average yields, which are influenced by local farming practices. While these discrepancies between the country and global levels are insightful, it provides a valuable opportunity for model improvement.

The simulated yield was compared with that of the GDHY dataset at the grid-cell level. The GDHY dataset is derived from census and remote sensing data, and may have introduced uncertainties into the evaluation results. Comparative studies with other soybean models and refining MATCRO-Soy on the basis of these findings will contribute to a more comprehensive understanding of its capabilities and limitations. Incorporating additional datasets will further enhance the MATCRO-Soy model's representation of real-world conditions. McCormick et al. (2021) suggested that integrating machine learning models will improve accuracy through the calibration process with numerous datasets. However, the use of mechanistic models embedded in MATCRO to simplify the process has proven valuable for understanding and predicting the impacts of environmental factors on agricultural systems. This model can be used to identify potential adaptation strategies, such as

changes in planting dates or the development of new crop varieties, to mitigate the adverse effects of climate change on soybean production. However, the application of this model at the field-scale requires high-quality data inputs and local parameter data.

7 Conclusions

We used MATCRO, which incorporates carbon assimilation modules based on C_3 photosynthesis of the Farquhar model, to simulate global soybean yield. The inputs were eco-physiological integrated gridded data related to climate, soil type, and nitrogen fertilization. Experimental datasets and information from previous studies were used to refine MATCRO-Soy so that it represents soybean growth under various environmental conditions. An evaluation of the global mean yields revealed a statistical correlation of 0.81 ($p < 0.001$) between the simulated yields and yields reported by FAOSTAT without subtracting the long-term yield trend. The correlation value was lower between simulated yields and detrended yield data. On the basis of comparisons of modelled and observed yields over a 34-year period (1981–2014), the correlation coefficients were 0.45 ($p < 0.050$) on the global scale and 0.52 ($p < 0.001$) for the top 10 soybean-producing countries. At the grid-cell level, the correlation between modelled and observed yields were significant in 66 % of grid cells. Therefore, the model successfully captured long-term trends and interannual variability, demonstrating its capacity to reflect the impacts of climate factors. Moreover, MATCRO-Soy also modelled reasonable photosynthetic processes at the site-scale, demonstrating its ability to represent temporal variations. This result highlights the model's reliability and adaptability as a tool for understanding soybean growth and yield dynamics.

While MATCRO-Soy presents a valuable framework for understanding the impacts of climate change on global soybean production, many localized factors that influence soybean yield resulting from shifts in climate (e.g., pests and diseases) can lead to discrepancies in yield prediction. This highlights the need for high-quality data inputs. The integration of CO_2 dynamics in MATCRO enhances crop response modelling because it includes the carbon fertilization effect. This warrants further investigation, along with analyses of the effects of other greenhouse gases. The model may benefit from further refinement, particularly in its treatment of temperature extremes, transpirable soil water, and nitrogen uptake during the photosynthesis process. Integrating MATCRO with other environmental models will enhance its applicability in agricultural management, although we emphasize the necessity for field-scale calibration to improve its reliability. MATCRO-Soy provides an opportunity to estimate changes in global soybean production under future land-use or climate change scenarios to address the complexities of climate interactions with agricultural systems. Overall, MATCRO-Soy has proven to be useful in understanding eco-physiological processes at the global, country, and grid-cell levels, providing valuable insights for agricultural management and climate change adaptation.

Code availability: This study used the model simulated by the source code of MATCRO (Yusara et al., 2025) archived at the following URL: https://doi.org/10.5281/zenodo.14881385.

Author Contributions: TK supervised this this study and acquired the funding. YM developed the model source code and supervised this study. AY adjusted the source code for the new parameterization, conducted analyses, and constructed figures. RB, LA, YW, SN, and EK provided the data for model calibration. AY prepared the original draft with further editing by YM, TK, LA, YT, KK, RB, EK, SN, and YW. KK supervised the statistical analysis of MSD and constructed figures. YM, TK, KK, YT, LA, RB, EK, SN, and YW reviewed it. The final manuscript was written by AY and approved by all authors.

Acknowledgements: The authors thanks all the contributors, and Maki Nakaguki and Marin Nagata in Hokkaido University for their assistance in using the model. We acknowledge the data provided by Inter-Sectoral Impact Model Intercomparison Project.

Competing interests: The authors declare that they have no conflict of interest.

Financial support: This work was supported by JST SPRING (Grant Number JPMJSP2119), and partially supported by grants from The Tojuro Iijima Foundation, National Key Research and Development Program of China (2022YFD2300901), the Environment Research and Technology Development Fund (JPMEERF21S12020 and JPMEERF20252002) of the Environmental Restoration and Conservation Agency (ERCA), and JSPS KAKENHI (Grant Number JP23H00351).

References

Adeboye, O. B., Schultz, B., Adeboye, A. P., Adekalu, K. O., and Osunbitan, J. A.: Application of the AquaCrop model in decision support for optimization of nitrogen fertilizer and water productivity of soybeans, Information Processing in Agriculture, 8, 419–436, https://doi.org/10.1016/j.inpa.2020.10.002, 2021.

Ai, Z. Hanasaki, N.: Simulation of crop yield using the global hydrological model H08 (crp.v1), Geosci Model Dev, 16, 3275–3290, https://doi.org/10.5194/gmd-16-3275-2023, 2023.

Ainsworth, E. A., Davey, P. A., Bernacchi, C. J., Dermody, O. C., Heaton, E. A., Moore, D. J., Morgan, P. B., Naidu, S. L., Ra, H. S. Y., Zhu, X. G., Curtis, P. S., and Long, S. P.: A meta-analysis of elevated [CO_2] effects on soybean (Glycine max) physiology, growth and yield, Glob Chang Biol, 8, 695–709, https://doi.org/10.1046/j.1365-2486.2002.00498.x, 2002.

Ainsworth, E. A., Rogers, A., Leakey, A. D. B., Heady, L. E., Gibon, Y., Stitt, M., and Schurr, U.: Does elevated atmospheric [CO2] alter diurnal C uptake and the balance of C and N metabolites in growing and fully expanded soybean leaves?, J Exp Bot, 58, 579–591, https://doi.org/10.1093/jxb/erl233, 2007.

Ainsworth, E. A., Serbin, S. P., Skoneczka, J. A., and Townsend, P. A.: Using leaf optical properties to detect ozone effects on foliar biochemistry, Photosynth Res, 119, 65–76, https://doi.org/10.1007/s11120-013-9837-y, 2014.

Anderson, W., Seager, R., Baethgen, W., and Cane, M.: Crop production variability in North and South America forced by life-cycles of the El Niño Southern Oscillation, Agric For Meteorol, 239, 151–165, https://doi.org/10.1016/j.agrformet.2017.03.008, 2017.

Ball, J. T.: An Analysis of Stomatal Conductance, Stanford University, CA, USA, 1988.

Basso, B., Ritchie, J. T., Pierce, F. J., Braga, R. P., and Jones, J. W.: Spatial validation of crop models for precision agriculture, Agric Syst, 68, 97–112, https://doi.org/10.1016/S0308-521X(00)00063-9, 2001.

Battisti, R., Sentelhas, P. C., and Boote, K. J.: Inter-comparison of performance of soybean crop simulation models and their ensemble in southern Brazil, Field Crops Res, 200, 28–37, https://doi.org/10.1016/j.fcr.2016.10.004, 2017.

Battisti, R., Sentelhas, P. C., Parker, P. S., Nendel, C., Câmara, G. M. D. S., Farias, J. R. B., and Basso, C. J.: Assessment of crop-management strategies to improve soybean resilience to climate change in Southern Brazil, Crop Pasture Sci, 69, 154–162, https://doi.org/10.1071/CP17293, 2018a.

Battisti, R., Sentelhas, P. C., and Boote, K. J.: Sensitivity and requirement of improvements of four soybean crop simulation models for climate change studies in Southern Brazil, Int J Biometeorol, 62, 823–832, https://doi.org/10.1007/s00484-017-1483-1, 2018b.

Bernacchi, C.J., Singsaas, E.L., Pimentel, C.A.R.L.O.S., Portis Jr, A.R. and Long, S.P.: Improved temperature response functions for models of Rubisco-limited photosynthesis. Plant, cell & environment, 24(2), pp.253-259, 2001.

Bernacchi, C. J., Morgan, P. B., Ort, D. R., and Long, S. P.: The growth of soybean under free air [CO2] enrichment (FACE) stimulates photosynthesis while decreasing in vivo Rubisco capacity, Planta, 220, 434–446, https://doi.org/10.1007/s00425-004-1320-8, 2005.

Bonan, G. B., Lawrence, P. J., Oleson, K. W., Levis, S., Jung, M., Reichstein, M., Lawrence, D. M., and Swenson, S. C.: Improving canopy processes in the Community Land Model version 4 (CLM4) using global flux fields empirically inferred from FLUXNET data, J Geophys Res, 116, https://doi.org/10.1029/2010jg001593, 2011.

Boote, K. J., Jones, J. W., Hoogenboom, G., and Pickering, N. B.: The CROPGRO model for grain legumes, 1983, 99–128, https://doi.org/10.1007/978-94-017-3624-4_6, 1998.

Boote, K. J., Jones, J. W., White, J. W., Asseng, S., and Lizaso, J. I.: Putting mechanisms into crop production models, Plant Cell Environ, 36, 1658–1672, https://doi.org/10.1111/pce.12119, 2013.

Bouman, B. A. M., Kropff, M. J., Tuong, T. P., Wopereis, M. C. S., ten Berge, H. F. M., and van Laar, H. H.: ORYZA2000 : modeling lowland rice, IRRI, 235 pp., 2001.

Büchner, M., Reyer, C. P. O.:: ISIMIP3b atmospheric composition input data (v1.1), ISIMIP Repository [data set], https://doi.org/10.48364/ISIMIP.482153.1, 2022.

Chen, C.-C. and Mccarl, B. A.: An investigation of the relationship between pesticide usage and climate change, ClimaticChange, 50, 475–487, 2001.

Collatz, G. J., Ball, J. T., Grivet, C., and Berry, J. A.: Physiological and environmental regulation of stomatal conductance, photosynthesis and transpiration: a model that includes a laminar boundary layer, Agric For Meteorol, 54, 107–136, https://doi.org/10.1016/0168-1923(91)90002-8, 1991.

Cucchi, M., Weedon, G. P., Amici, A., Bellouin, N., Lange, S., Müller Schmied, H., Hersbach, H., and Buontempo, C.: WFDE5: bias-adjusted ERA5 reanalysis data for impact studies, Earth Syst. Sci. Data, 12, 2097–2120, https://doi.org/10.5194/essd-12-2097-2020, 2020.

Cuddington, K., Fortin, M. J., Gerber, L. R., Hastings, A., Liebhold, A., O'connor, M., and Ray, C.: Process-based models are required to manage ecological systems in a changing world, Ecosphere, 4, https://doi.org/10.1890/ES12-00178.1, 2013.

Dai, Y., Dickinson, R. E., and Wang, Y.-P.: A Two-Big-Leaf Model for Canopy Temperature, Photosynthesis, and Stomatal Conductance, J Clim, 17, 2281–2299, 2004.

De Pury, D. G. G. and Farquhar, G. D.: Simple scaling of photosynthesis from leaves to canopies without the errors of big-leaf models, Plant Cell Environ, 20, 537–557, https://doi.org/10.1111/j.1365-3040.1997.00094.x, 1997.

De Vries, F. W. T. P. de., Jansen, D. M., Berge, H. F. M. ten, and Bakema, A.: Simulation of ecophysiological processes of growth in several annual crops, Pudoc, 271 pp., 1989.

Egli, D. B., Bruening, W. P.: Water stress, photosynthesis, seed sucrose levels and seed growth in soybean, The Journal of Agricultural Science, 142(1), 1-8, 2004.

Elliott, J., Müller, C., Deryng, D., Chryssanthacopoulos, J., Boote, K. J., Büchner, M., Foster, I., Glotter, M., Heinke, J., Iizumi, T., Izaurralde, R. C., Mueller, N. D., Ray, D. K., Rosenzweig, C., Ruane, A. C., and Sheffield, J.: The Global Gridded Crop Model Intercomparison: Data and modeling protocols for Phase 1 (v1.0), Geosci Model Dev, 8, 261–277, https://doi.org/10.5194/gmd-8-261-2015, 2015.

Farquhar, G. D., Von Caemmerer, S., and Berry, J. A.: A Biochemical Model of Photosynthetic CO_2 Assimilation in Leaves of C 3 Species, Planta, 78–90 pp., 1980.

Fehr, W. R. and Caviness, C. E.: "Stages of soybean development" (1977). Special Report. 80., Iowa State University, 1–12, 1977.

Fisher, J. B., S. Sitch, Y. Malhi, R. A. Fisher, C. Huntingford, and S-Y. Tan.: Carbon cost of plant nitrogen acquisition: A mechanistic, globally applicable model of plant nitrogen uptake, retranslocation, and fixation, Global Biogeochemical Cycles, 24, GB1014, 2010.

Fodor, N., Challinor, A., Droutsas, I., Ramirez-Villegas, J., Zabel, F., Koehler, A. K., and Foyer, C. H.: Integrating plant science and crop modeling: assessment of the impact of climate change on soybean and maize production, Plant and Cell Physiology, 58, https://doi.org/10.1093/pcp/pcx141, 2017.

García-Tejero, I. F., Durán-Zuazo, V. H., Muriel-Fernández, J. L., and Rodríguez-Pleguezuelo, C. R.: Water and Sustainable Agriculture, Springer Netherlands, Dordrecht, https://doi.org/10.1007/978-94-007-2091-6, 2011.

Gauch, H. G., Hwang, J. T. G., Fick, G. W., Gauch, H. G., and Fick, G. W.: Model Evaluation by Comparison of Model-Based Predictions and Measured Values, Agron J, 95, 1442–1446, 2003.

Hoogenboom, G.: Contribution of agrometeorology to the simulation of crop production and its applications, Agric For Meteorol, 103, 137–157, https://doi.org/10.1016/S0168-1923(00)00108-8, 2000.

Hurtt, G. C., Chini, L., Sahajpal, R., Frolking, S., Bodirsky, B. L., Calvin, K., Doelman, J. C., Fisk, J., Fujimori, S., Goldewijk, K. K., Hasegawa, T., Havlik, P., Heinimann, A., Humpenöder, F., Jungclaus, J., Kaplan, J. O., Kennedy, J., Krisztin, T., Lawrence, D., Lawrence, P., Ma, L., Mertz, O., Pongratz, J., Popp, A., Poulter, B., Riahi, K., Shevliakova, E., Stehfest, E., Thornton, P., Tubiello, F. N., van Vuuren, D. P., and Zhang, X.: Harmonization of global land use change and management for the period 850-2100 (LUH2) for CMIP6, Geosci Model Dev, 13, 5425–5464, https://doi.org/10.5194/gmd-13-5425-2020, 2020.

Iizumi, T.: Global dataset of historical yields v1.2 and v1.3 aligned version, https://doi.org/10.1594/PANGAEA.909132, 2019.

Iizumi, T., Luo, J. J., Challinor, A. J., Sakurai, G., Yokozawa, M., Sakuma, H., Brown, M. E., and Yamagata, T.: Impacts of El Niño Southern Oscillation on the global yields of major crops, Nat Commun, 5, https://doi.org/10.1038/ncomms4712, 2014.

Iizumi, T., Sakai, T. The global dataset of historical yields for major crops 1981-2016. Scientific Data 7, 97, https://doi.org/10.1038/s41597-020-0433-7, 2020.

Irmak, S., Odhiambo, L. O., Specht, J. E., and Djaman, K.: Hourly and daily single and basal evapotranspiration crop coefficients as a function of growing degree days, days after emergence, leaf area index, fractional green canopy cover, and plant phenology for soybean, Trans ASABE, 56, 1785–1803, 2013.

Jägermeyr, J., Müller, C., Ruane, A. C., Elliott, J., Balkovic, J., Castillo, O., Faye, B., Foster, I., Folberth, C., Franke, J. A., Fuchs, K., Guarin, J. R., Heinke, J., Hoogenboom, G., Iizumi, T., Jain, A. K., Kelly, D., Khabarov, N., Lange, S., Lin, T. S., Liu, W., Mialyk, O., Minoli, S., Moyer, E. J., Okada, M., Phillips, M., Porter, C., Rabin, S. S., Scheer, C., Schneider, J. M., Schyns, J. F., Skalsky, R., Smerald, A., Stella, T., Stephens, H., Webber, H., Zabel, F., and Rosenzweig, C.: Climate impacts on global agriculture emerge earlier in new generation of climate and crop models, Nat Food, 2, 873–885, https://doi.org/10.1038/s43016-021-00400-y, 2021.

Jin, Z., Ainsworth, E. A., Leakey, A. D. B., and Lobell, D. B.: Increasing drought and diminishing benefits of elevated carbon dioxide for soybean yields across the US Midwest, Glob Chang Biol, 24, e522–e533, https://doi.org/10.1111/gcb.13946, 2018.

Jones, J. W., Antle, J. M., Basso, B., Boote, K. J., Conant, R. T., Foster, I., Godfray, H. C. J., Herrero, M., Howitt, R.E., Janssen, S., Keating, B.A., Munoz-Carpena, R., Porter, C. H., Rosenzweig, C. ,Wheeler, T. R.: Brief history of agricultural systems modeling. Agricultural systems, 155, 240-254, 2017.

Kawasaki, Y., Katsura, K., and Shiraiwa, T.: Yield and dry matter dynamics of vegetative and reproductive organs in Japanese and US soybean cultivars, Plant Prod Sci, 21, 349–357, https://doi.org/10.1080/1343943X.2018.1512874, 2018.

Kim, H: Global Soil Wetness Project Phase 3 Atmospheric Boundary Conditions (Experiment 1), Data Integration and Analysis System (DIAS) [data set], https://doi.org/10.20783/DIAS.501, 2017.

Kinose, Y., Masutomi, Y., Shiotsu, F., Hayashi, K., Ogawada, D., Gomez-Garcia, M., Matsumura, A., Takahashi, K., and Fukushi, K.: Impact assessment of climate change on the major rice cultivar ciherang in Indonesia, Journal of Agricultural Meteorology, 76, 19–28, https://doi.org/10.2480/agrmet.D-19-00045, 2020.

Kobayashi, K. and Salam, M. U.: Comparing simulated and measured values using mean squared deviation and its Components, Agron J, 92, 345–352, 2000.

Kothari, K., Battisti, R., Boote, K. J., Archontoulis, S. V., Confalone, A., Constantin, J., Cuadra, S. V., Debaeke, P., Faye, B., Grant, B., Hoogenboom, G., Jing, Q., van der Laan, M., Macena da Silva, F. A., Marin, F. R., Nehbandani, A., Nendel, C., Purcell, L. C., Qian, B., Ruane, A. C., Schoving, C., Silva, E. H. F. M., Smith, W., Soltani, A., Srivastava, A., Vieira, N. A., Slone, S., and Salmerón, M.: Are soybean models ready for climate change food impact assessments?, European Journal of Agronomy, 135, https://doi.org/10.1016/j.eja.2022.126482, 2022.

Kumagai, E.: Effect of early sowing on growth and yield of determinate and indeterminate soybean (Glycine max (L.) Merr) cultivars in a cool region of northern Japan, Journal of Agricultural Meteorology, 74, 18–28, https://doi.org/10.2480/agrmet.D-17-00009, 2018.

Kumagai, E.: Agronomic responses of soybean cultivars to narrow intra-row spacing in a cool region of northern Japan, Plant Prod Sci, 24, 29–40, https://doi.org/10.1080/1343943X.2020.1816137, 2021.

La Menza, N. C., Arkebauer, T. J., Lindquist, J. L., Monzon, J. P., Knops, J. M. H., Graef, G., Scoby, D., Howard, R., Rees, J., Specht, J. E., and Grassini, P.: Decoupling between leaf nitrogen and radiation-use efficiency in vegetative and early reproductive stages in high-yielding soybean, J Exp Bot, https://doi.org/10.1093/jxb/erac408, 2022.

Lange, S.: Trend-preserving bias adjustment and statistical downscaling with ISIMIP3BASD (v1.0), Geosci. Model Dev., 12, 3055–3070, https://doi.org/10.5194/gmd-12-3055-2019, 2019.

Lange, S., Menz, C., Gleixner, S., Cucchi, M., Weedon, G.P., Amici, A., Bellouin, N., Müller-Schmied, H., Hersbach, H., Buontempo, C., Cagnazzo, C: WFDE5 over land merged with ERA5 over the ocean (W5E5 v2.0), ISIMIP Repository [data set]. https://doi.org/10.48364/ISIMIP.342217, 2021.

Leung, F., Williams, K., Sitch, S., Tai, A. P. K., Wiltshire, A., Gornall, J., Ainsworth, E. A., Arkebauer, T., and Scoby, D.: Calibrating soybean parameters in JULES 5.0 from the US-Ne2/3 FLUXNET sites and the SoyFACE-O_3 experiment, Geosci Model Dev, 13, 6201–6213, https://doi.org/10.5194/gmd-13-6201-2020, 2020.

Long, S. P., Ainsworth, E. A., Leakey, A. D. B., and Morgan, P. B.: Global food insecurity. Treatment of major food crops with elevated carbon dioxide or ozone under large-scale fully open-air conditions suggests recent models may have overestimated future yields, in: Philosophical Transactions of the Royal Society B: Biological Sciences, 2011–2020, https://doi.org/10.1098/rstb.2005.1749, 2005.

Ma, J., Olin, S., Anthoni, P., Rabin, S. S., Bayer, A. D., Nyawira, S. S., and Arneth, A.: Modeling symbiotic biological nitrogen fixation in grain legumes globally with LPJ-GUESS, Geosci Model Dev, 15, 815–839, https://doi.org/10.5194/gmd-15-815-2022, 2022.

Marin, F. R., Ribeiro, R. V., and Marchiori, P. E. R.: How can crop modeling and plant physiology help to understand the plant responses to climate change? A case study with sugarcane, Theor Exp Plant Physiol, 26, 49–63, https://doi.org/10.1007/s40626-014-0006-2, 2014.

Marin, F. R., Zanon, A. J., Monzon, J. P., Andrade, J. F., Silva, E. H. F. M., Richter, G. L., Antolin, L. A. S., Ribeiro, B. S. M. R., Ribas, G. G., Battisti, R., Heinemann, A. B., and Grassini, P.: Protecting the Amazon forest and reducing global warming via agricultural intensification, Nat Sustain, 5, 1018–1026, https://doi.org/10.1038/s41893-022-00968-8, 2022.

Masutomi, Y., Ono, K., Mano, M., Maruyama, A., and Miyata, A.: A land surface model combined with a crop growth model for paddy rice (MATCRO-Rice v. 1) - Part 1: Model description, Geosci Model Dev, 9, 4133–4154, https://doi.org/10.5194/gmd-9-4133-2016, 2016a.

Masutomi, Y., Ono, K., Takimoto, T., Mano, M., Maruyama, A., and Miyata, A.: A land surface model combined with a crop growth model for paddy rice (MATCRO-Rice v. 1) - Part 2: Model validation, Geosci Model Dev, 9, 4155–4167, https://doi.org/10.5194/gmd-9-4155-2016, 2016b.

Masutomi, Y., Kinose, Y., Takimoto, T., Yonekura, T., Oue, H., and Kobayashi, K.: Ozone changes the linear relationship between photosynthesis and stomatal conductance and decreases water use efficiency in rice, Science of the Total Environment, 655, 1009–1016, https://doi.org/10.1016/j.scitotenv.2018.11.132, 2019.

McCormick, R. F., Truong, S. K., Rotundo, J., Gaspar, A. P., Kyle, D., Van Eeuwijk, F., and Messina, C. D.: Intercontinental prediction of soybean phenology via hybrid ensemble of knowledge-based and data-driven models, In Silico Plants, 3, https://doi.org/10.1093/insilicoplants/diab004, 2021.

Morgan, P. B., Bollero, G. A., Nelson, R. L., Dohleman, F. G., and Long, S. P.: Smaller than predicted increase in aboveground net primary production and yield of, Glob Chang Biol, 11, 1856–1865, https://doi.org/10.1111/j.1365-2486.2005.01017.x, 2005.

Müller, C., Elliott, J., Chryssanthacopoulos, J., Arneth, A., Balkovic, J., Ciais, P., Deryng, D., Folberth, C., Glotter, M., Hoek, S., Iizumi, T., Izaurralde, R. C., Jones, C., Khabarov, N., Lawrence, P., Liu, W., Olin, S., Pugh, T. A. M., Ray, D. K., Reddy, A., Rosenzweig, C., Ruane, A. C., Sakurai, G., Schmid, E., Skalsky, R., Song, C. X., Wang, X., De Wit, A., and Yang, H.: Global gridded crop model evaluation: Benchmarking, skills, deficiencies and implications, Geosci Model Dev, 10, 1403–1422, https://doi.org/10.5194/gmd-10-1403-2017, 2017.

Nakano, S., Homma, K., and Shiraiwa, T.: Modeling biomass and yield production based on nitrogen accumulation in soybean grown in upland fields converted from paddy fields in Japan, Plant Prod Sci, 24, 440–453, https://doi.org/10.1080/1343943X.2021.1881409, 2021.

Portmann, F. T., Siebert, S., and Döll, P.: MIRCA2000-Global monthly irrigated and rainfed crop areas around the year 2000: A new high-resolution data set for agricultural and hydrological modelling, Global Biogeochemical Cycles, 24(1), http://doi.org/10.1029/2008GB003435, 2010.

Porwollik, V., Müller, C., Elliott, J., Chryssanthacopoulos, J., Iizumi, T., Ray, D. K., Ruane, A. C., Arneth, A., Balkovič, J., Ciais, P., Deryng, D., Folberth, C., Izaurralde, R. C., Jones, C. D., Khabarov, N., Lawrence, P. J., Liu, W., Pugh, T. A. M., Reddy, A., Sakurai, G., Schmid, E., Wang, X., de Wit, A., and Wu, X.: Spatial and temporal uncertainty of crop yield aggregations, European Journal of Agronomy, 88, 10–21, https://doi.org/10.1016/j.eja.2016.08.006, 2017.

Qiang, B., Zhou, W., Zhong, X., Fu, C., Cao, L., Zhang, Y., and Jin, X.: Effect of nitrogen application levels on the photosynthetic nitrogen distribution and use efficiency in leaves of soybean seedlings, https://doi.org/10.21203/rs.3.rs-2060928/v1, 2022.

Ray, D. K., Gerber, J. S., Macdonald, G. K., and West, P. C.: Climate variation explains a third of global crop yield variability, Nat Commun, 6, https://doi.org/10.1038/ncomms6989, 2015.

Ray, J. D. and Sinclair, T. R.: The effect of pot size on growth and transpiration of maize and soybean during water deficit stress, J Exp Bot, 49, 1381–1386, https://doi.org/10.1093/jxb/49.325.1381, 1998.

Ruane, A. C., Rosenzweig, C., Asseng, S., Boote, K. J., Elliott, J., Ewert, F., Jones, J. W., Martre, P., McDermid, S. P., Müller, C., Snyder, A., and Thorburn, P. J.: An AgMIP framework for improved agricultural representation in integrated assessment models, Environmental Research Letters, 12, https://doi.org/10.1088/1748-9326/aa8da6, 2017.

Rohatgi, A. WebPlotDigitizer (Version 5.2) [software], https://automeris.io, 2023.

Sage, R. F.: Variation in the kcat of Rubisco in C_3 and C_4 plants and some implications for photosynthetic performance at high and low temperature, J Exp Bot, 53, 609–620, 2002.

Sakurai, G., Iizumi, T., Nishimori, M., and Yokozawa, M.: How much has the increase in atmospheric CO_2 directly affected past soybean production?, Sci Rep, 4, https://doi.org/10.1038/srep04978, 2014.

Santachiara, G., Salvagiotti, F., and Rotundo, J. L.: Nutritional and environmental effects on biological nitrogen fixation in soybean: A meta-analysis, Field Crops Res, 240, 106–115, https://doi.org/10.1016/j.fcr.2019.05.006, 2019.

Scafaro, A. P., Posch, B. C., Evans, J. R., Farquhar, G. D., and Atkin, O. K.: Rubisco deactivation and chloroplast electron transport rates co-limit photosynthesis above optimal leaf temperature in terrestrial plants, Nat Commun, 14, https://doi.org/10.1038/s41467-023-38496-4, 2023.

Sinclair, T. R.: Water and nitrogen limitations in soybean grain production I. Model development, Field Crops Res, 15, 125–141, https://doi.org/10.1016/0378-4290(86)90082-1, 1986.

Slattery, R. A., Vanloocke, A., Bernacchi, C. J., Zhu, X. G., and Ort, D. R.: Photosynthesis, light use efficiency, and yield of reduced-chlorophyll soybean mutants in field conditions, Front Plant Sci, 8, https://doi.org/10.3389/fpls.2017.00549, 2017.

Smith, B., Wärlind, D., Arneth, A., Hickler, T., Leadley, P., Siltberg, J., and Zaehle, S.: Implications of incorporating N cycling and N limitations on primary production in an individual-based dynamic vegetation model, Biogeosciences, 11, 2027–2054, https://doi.org/10.5194/bg-11-2027-2014, 2014.

Stöckle, C. O. and Kemanian, A. R.: Can crop models identify critical gaps in genetics, environment, and management interactions?, Front. Plant Sci., 11, https://doi.org/10.3389/fpls.2020.00737, 2020.

Takata, K., Emori, S., and Watanabe, T.: Development of the minimal advanced treatments of surface interaction and runoff, Glob Planet Change, 38, 209–222, https://doi.org/10.1016/S0921-8181(03)00030-4, 2003.

Thies, J. E., Singleton, P. W., and Bohlool, B. B.: Phenology, growth, and yield of field-grown soybean and bush bean as a function of varying modes of N nutrition, Soil Biol Biochem, 27, 575–583, https://doi.org/10.1016/0038-0717(95)98634-Z, 1995.

Thompson, J. A., Schweitzer, E., and Nelson, R. L.: Association of specific leaf weight, an estimate of chlorophyll, and chlorophyll concentration with apparent photosynthesis in soybean, Phowsynthesis Research, Kluwer Academic Publishers, 1–10 pp., 1996.

Volkholz, J., Müller, C.: ISIMIP3 soil input data (v1.0), ISIMIP Repository [data set], https://doi.org/10.48364/ISIMIP.942125, 2020.

Volkholz, J., Sebastian Ostberg, S.: ISIMIP3a N-fertilizer input data (v1.2), ISIMIP Repository [data set], https://doi.org/10.48364/ISIMIP.311496.2, 2022.

Wirth, S. B., Braun, J., Heinke, J., Ostberg, S., Rolinski, S., Schaphoff, S., Stenzel, F., Von Bloh, W., and Müller, C.: Biological nitrogen fixation of natural and agricultural vegetation simulated with LPJmL 5.7.9, https://doi.org/10.5194/egusphere-2023-2946, 2024.

Wu, X., Vuichard, N., Ciais, P., Viovy, N., De Noblet-Ducoudré, N., Wang, X., Magliulo, V., Wattenbach, M., Vitale, L., Di Tommasi, P., Moors, E. J., Jans, W., Elbers, J., Ceschia, E., Tallec, T., Bernhofer, C., Grünwald, T., Moureaux, C., Manise, T., Ligne, A., Cellier, P., Loubet, B., Larmanou, E., and Ripoche, D.: ORCHIDEE-CROP (v0), a new process-based agro-land surface model: Model description and evaluation over Europe, Geosci Model Dev, 9, 857–873, https://doi.org/10.5194/gmd-9-857-2016, 2016.

Wu, Y., Wang, E., He, D., Liu, X., Archontoulis, S. V., Huth, N. I., Zhao, Z., Gong, W., and Yang, W.: Combine observational data and modelling to quantify cultivar differences of soybean, European Journal of Agronomy, 111, https://doi.org/10.1016/j.eja.2019.125940, 2019.

Xu, Q., Liu, X., Zhang, C., Du, W., Guan, Y., and Yang, W.: Insights into soybean with high photosynthetic efficiency, Adv Bot Res, 102, 121–151, https://doi.org/10.1016/BS.ABR.2022.02.019, 2022.

Yin, X.: Improving ecophysiological simulation models to predict the impact of elevated atmospheric CO2 concentration on crop productivity, https://doi.org/10.1093/aob/mct016, August 2013.

Yusara, A., Kato, T., Ainsworth, E., Rafael, B., Kumagai, E., Satoshi, N., Wu, Y., Tsusumi-Morita, Y., Kobayashi, K., Masutomi, Y.: An eco-physiological process-based model for soy yield (MATCRO-Soy v.1), Zenodo [code], https://doi.org/10.5281/zenodo.14881385, 2025.

Zaehle, S. Friend, A. D.: Carbon and nitrogen cycle dynamics in the O-CN land surface model: 1. Model description, site-scale evaluation, and sensitivity to parameter estimates, Global Biogeochem Cycles, 24, https://doi.org/10.1029/2009GB003521, 2010.

Zheng, Y., Li, F., Hao, L., Yu, J., Guo, L., Zhou, H., Ma, C., Zhang, X., and Xu, M.: Elevated CO_2 concentration induces photosynthetic down-regulation with changes in leaf structure, non-structural carbohydrates and nitrogen content of soybean, BMC Plant Biol, 19, https://doi.org/10.1186/s12870-019-1788-9, 2019.