Agricultural and Forest Meteorology 365 (2025) 110463

Contents lists available at ScienceDirect

Agricultural and Forest Meteorology

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

Inter-comparison of soybean models for the simulation of evapotranspiration in a humid continental climate

Evandro H. Figueiredo Moura da Silva[a], Kritika Kothari[b,c], Elizabeth Pattey[d], Rafael Battisti[e], Kenneth J. Boote[f], Sotirios V. Archontoulis[g], Santiago Vianna Cuadra[h], Babacar Faye[i], Brian Grant[d], Gerrit Hoogenboom[f,j], Qi Jing[d], Fábio R. Marin[k], Claas Nendel[l,m], Budong Qian[d], Ward Smith[d], Amit Kumar Srivastava[m,n], Kelly R. Thorp[o], Nilson A. Vieira Junior[p], Montserrat Salmerón[b,*]

[a] University of São Paulo, College of Agriculture "Luiz de Queiroz", Department of Soil Science, Piracicaba, SP, Brazil
[b] University of Kentucky, Department of Plant and Soil Sciences, Lexington, KY, United States
[c] Indian Institute of Technology Roorkee, Department of Water Resources Development and Management, Roorkee, Uttarakhand, India
[d] Agriculture and Agri-Food Canada, Ottawa Research and Development Centre, Ottawa, Canada
[e] Universidade Federal de Goiás, Escola de Agronomia, Goiânia, GO, Brazil
[f] University of Florida, Department of Agricultural and Biological Engineering, Gainesville, FL, United States
[g] Iowa State University, Department of Agronomy, Ames, IA, United States
[h] Brazilian Agricultural Research Corporation, Brasília, Brazil
[i] Université du Sine Saloum El-Hadj Ibrahima NIASS, 55 Kaolack, Senegal
[j] University of Florida, Global Food Systems Institute, Gainesville, FL, United States
[k] University of São Paulo, College of Agriculture "Luiz de Queiroz", Department of Biosystems Engineering, Piracicaba, SP, Brazil
[l] Institute of Biochemistry and Biology, University of Potsdam, Potsdam, Germany
[m] Leibniz Centre for Agricultural Landscape Research (ZALF), Müncheberg, Germany
[n] University of Bonn, Institute of Crop Science and Resource Conservation, Bonn, Germany
[o] USDA-ARS, Grassland, Soil, and Water Research Laboratory, Temple, TX, United States
[p] Wageningen University & Research, Wageningen, the Netherlands, 6708 PB

ARTICLE INFO

Keywords:
Crop transpiration
Soil water evaporation
Priestley-Taylor
FAO-56 Penman-Monteith
Eddy covariance
Water use

ABSTRACT

Accurate simulation of evapotranspiration (ET) with crop models is essential for improving agricultural water management and yield forecasting. Few studies have evaluated multiple soybean [*Glycine* max (L.) Merr.] models for simulating ET under conditions of low evaporative demand that is characteristic for a warm-summer humid continental climate. Six soybean crop models, encompassing 15 different modeling approaches, were evaluated for ET simulation and compared against eddy covariance data collected over five growing seasons in Ottawa, Canada. Models were first calibrated with phenology, in-season growth, and yield data, followed by calibration with measured ET and soil water content (SWC) data during the second step. After initial calibration, simulated daily ET was higher on average than measured ET, particularly during full canopy cover (normalized bias, nBias = 17.1 to 49.2% depending on the model). Following the second calibration, simulated daily ET was closer to measured values, but bias remained (nBias = 5.9 to 52.1% during full canopy). The ensemble median reduced uncertainty in the simulation of daily ET compared to most models, but DNDC remained the top-ranking model (nRMSE = 0.7 mm d^{-1}, nBias = 11.2%). The MONICA model was most accurate simulating cumulative ET (RMSE = 39.9 mm, nBias = 11.3%), whereas the CROPGRO models excelled simulating SWC (RMSE= 0.04 to 0.05 m^3 m^{-3}, nBias = 0.10 to 0.9% depending on soil depth). This study was instrumental in evaluating the best ET methodologies and parameters in soybean models. However, there was bias across the models compared to

Abbreviations: AGB, total aboveground biomass; ASCE, American Society of Civil Engineers; D-statistic, index of agreement; E, soil water evaporation; Ep, potential soil water evaporation; Es, simulated soil water evaporation; ET, evapotranspiration; ETo, reference evapotranspiration with grass as the reference surface (short crop); ETp, potential evapotranspiration; ETr, reference evapotranspiration with alfalfa as the reference surface (tall crop); ETs, simulated evapotranspiration; LAI, leaf area index; LC, lack of correlation; nBias, normalized bias; MSE, mean square error; RMSE, root mean square error; nRMSE, normalized RMSE; NU, non-unity of slope; SB, standard bias; SWC, soil water content; T, Plant transpiration; Ts, simulated plant transpiration; Tp, potential plant transpiration.

* Corresponding author.
E-mail address: msalmeron@uky.edu (M. Salmerón).

https://doi.org/10.1016/j.agrformet.2025.110463
Received 21 August 2024; Received in revised form 17 February 2025; Accepted 18 February 2025
Available online 1 March 2025

measured eddy covariance ET in a humid environment. The results reveal the need to further investigate possible biases in ET estimates by eddy covariance over soybean canopies, and to review the role of night-time dew contributions to ET in process-based models.

1. Introduction

Soybean, a primary protein source for both human consumption and animal feed, plays a crucial role in meeting global food and feed demand (Westhoek et al., 2011; Wu et al., 2014). Major soybean production areas in the world include regions with a humid continental climate, such as the U.S. Corn Belt, northeastern China, and eastern Canada (Wang et al., 2020; USDA, 2022; SOYCANADA, 2022). These production areas require continuous advances in soil, crop, and water management practices to ensure sustainable production.

Water scarcity is considered one of the biggest challenges to meeting the increasing food demand, impairing crop metabolic processes, biological nitrogen fixation, and ultimately seed yield (Purcell and King, 1996; Figueiredo Moura da Silva et al., 2023). To mitigate the effects of water stress and drought, irrigation has proliferated in the water-limited agricultural production centers of the world (Siebert et al., 2005; Puy et al., 2020, 2021; Din et al., 2022). Currently, arable lands account for 60% of global freshwater withdrawals (Wu et al., 2022), and the projected increase in irrigated areas will further escalate freshwater demand.

An accurate estimation of water use is crucial for determining crop water requirements and developing efficient water management strategies (Hoogeveen et al., 2009; Chartres and Noble, 2015; Ndehedehe, 2022). Evapotranspiration (ET) represents the total water use from agricultural land in the form of soil water evaporation (E) and crop transpiration (T). There are several methods to obtain indirect measurements of ET, including the Bowen Ratio-Energy Balance (Bowen, 1926), eddy covariance (Mizutani et al., 1997), and weighing lysimeters (Allen and Fisher, 1991). Although these methods can be precise, they are labor-intensive, site-specific, and expensive. Process-based crop eco-physiological models, after accurate calibration, can be robust tools for estimating ET and exploring management strategies that increase the sustainability of water use in agriculture (Keating and Thorburn, 2018; Figueiredo Moura da Silva et al., 2021, 2022). The conceptual underpinning for crop models is the simulation of the dynamic effects of environmental variables on daily ET and drought stress, crop development, growth, and yield (Hoogenboom, 2000; Boote et al., 2008; White et al., 2011; Rosenzweig et al., 2014).

Process-based crop models use various methods for estimating ET, as described in a study evaluating 29 maize (*Zea mays* L.) models by Kimball et al. (2019). The different methods to compute ET can introduce significant uncertainty in the simulation of crop water demand and yield. Therefore, evaluating approaches to estimate ET in crop models with measured ET data is essential to reduce uncertainty before model application. For maize, Kimball et al. (2019, 2023) compared multiple models against ET data from eddy covariance and weighing lysimeters under a semi-arid and a hot-summer humid continental climate. Their results revealed that the ensemble median generally outperformed individual models for simulation of daily ET, with some exceptions depending on the calibration phase (Kimball et al., 2019, 2023). Thus, relying on models that employ a wide range of approaches for simulation of ET might be necessary to reduce uncertainty in ET simulations, particularly for locations with limited ET and soil moisture data for calibration.

Multi-model studies can also identify approaches within models that improve simulation of ET. For example, Sau et al. (2004) using neutron probe measurements in a humid oceanic climate in northwestern Spain, and Figueiredo Moura da Silva et al. (2022) using Bowen Ratio-Energy Balance in a tropical environment in Brazil, evaluated ET methods within the DSSAT-CSM-CROPGRO model. Both studies concluded that the FAO-56 Penman-Monteith method (Allen et al., 1998) provided a better fit than the Priestley–Taylor method (Priestley and Taylor, 1972) when compared with measured ET or soil water content data. Similarly, Thorp et al. (2020) found that the FAO-56 Penman-Monteith method outperformed the Priestley–Taylor method within DSSAT-CSM-CROPGRO for cotton ET simulations in Bushland, TX, using lysimeter data (Priestley and Taylor, 1972). For soybean, the CROPGRO model within DSSAT-CSM currently includes 17 different combinations of potential ET and soil evaporation methods, including an energy balance version (Cuadra et al., 2021) and potential ET based on ASCE approaches (DeJonge and Thorp, 2017). Thus, it is crucial to conduct a coordinated multi-model evaluation of ET methods within DSSAT-CSM-CROPGRO-Soybean and against the same experimental dataset to facilitate selection of ET approaches for model users.

Prior studies evaluating crop models for the simulation of ET with measured data are relatively limited for soybean, particularly under a warm-summer humid continental climate. This climate is characterized by a relatively low evapotranspiration demand, and less likelihood of drought stress compared to environmental conditions in most prior ET studies (Kimball et al. 2019, 2023; Sau et al. 2004; Figueiredo Moura da Silva et al., 2022; Thorp et al., 2020). Of interest, Sansoulet et al. (2014) found a systematic overestimation of ET by the wheat models that they evaluated against eddy covariance measurements in a warm-summer humid continental climate in eastern Canada. Overall, the evaluation of crop models with experimental data that includes ET measurements remains a major bottleneck for reducing model uncertainty for simulating water dynamics, particularly for a major crop like soybean for a warm-summer humid continental climate.

A coordinated evaluation of multiple models against a common dataset and under blind conditions provides additional advantages to quantifying current uncertainty in model simulations and identifying the best approaches in crop models to increase their robustness. Groups within the Agricultural Model Intercomparison and Improvement Project (AgMIP) were established to quantify and reduce uncertainties by conducting multi-model ensemble studies, in which the variability between models is an indication of the structural uncertainty (Rosenzweig et al., 2013, 2021). Previous multi-model studies within AgMIP have focused on evaluating models against experimental data for wheat (Asseng et al., 2013; Cammarano et al., 2016; Wang et al., 2017; Guarin and Asseng, 2022), maize (Bassu et al., 2014; Kimball et al., 2019, 2023; Yasin et al., 2022), rice (Li et al., 2015), sugarcane (Marin et al., 2015), potato (Fleisher et al., 2017), and soybean (Battisti et al., 2017; Kothari et al., 2022). However, only Kimball et al. (2019, 2023) evaluated different maize models with measured ET data from eddy-covariance flux towers or weighing lysimeters. Thus, there is still a need to evaluate models for other crops for the simulation of ET dynamics as compared to measured data.

Accurate ET simulations are essential for using models to develop sustainable management practices for cropping systems. For this purpose, evaluating crop models for the simulation of ET with measured data is paramount. In addition, a coordinated multi-model evaluation with a common dataset and using different methods to compute ET can provide opportunities to identify the best methods to simulate ET for further model improvement. The primary goal of this study was to evaluate the performance of different soybean models in simulating ET, soil water content (SWC), and crop growth for a warm-summer humid continental climate.. This research represents the first comprehensive assessment of multiple soybean crop models under such climatic conditions. Specific objectives of this study were to (i) compare 15 flavors of models (based on six crop models with two flavors of APSIM and nine flavors of DSSAT-CSM-CROPGRO) for the simulation of ET and SWC after calibrating models with crop growth data (step 1), and after calibration with measured ET and SWC data (step 2), and (ii) quantify the applicability of a model ensemble to reduce uncertainty in the simulation of ET and SWC, and (iii) discuss opportunities for model improvement. Models were evaluated against detailed eddy covariance flux measurements of ET, in-season growth data, and time domain reflectometry measurements of SWC from field experiments conducted in Canada during 5 years (with 6 environments as the 2016 year had two sowing dates).

2. Materials and methods

2.1. Field experimental data

The ET and ancillary meteorological datasets were measured using eddy flux and meteorological towers installed in large fields, with a minimum fetch of 200 m (Pattey et al., 2006). The fields were located in the Greenbelt Experimental Farm (GEF) during the 1997 and 1999 seasons and the adjacent farm of the Canadian Food Inspection Agency Fallowfield Laboratory (CFIA) during the 2008, 2016, and 2019 seasons in Ottawa, Ontario, Canada (45.3° N, 75.8° W, 91 m a.s.l.). According to the Koeppen (1948) climate classification, the site is Dfb (Warm Summer Humid Continental Climate), with a mean annual temperature of 6.4 °C and annual total precipitation of 943 mm (1981–2010 normal, Environment and Climate Change Canada, 2022).

Daily ET was estimated half-hourly using the eddy covariance method, and the data were screened and aggregated to daily and seasonal ET following the method of Pattey et al. (2001). The flux measuring systems were installed at 2-m height above the plant canopies to record the following measurements: (i) wind speed and sonic temperature with three-dimensional ultrasonic anemometer-thermometers (1997–2008: DAT-310, Kaijo-Denki Co., Tokyo, Japan; 2016–19: CSAT-3, CSI, Logan Utah) and (ii) H_2O concentration with closed-path infrared gas analyzers (1997–99: LI-6262, 2008–19: LI-7000, Li-Cor Inc., Lincoln, NE). Weather data were measured in situ, including solar radiation (MJ m^{-2} d^{-1}), daily maximum and minimum temperature (°C); dewpoint temperature (°C), daily total rainfall (mm); and daily average wind speed (m s^{-1}) at a height of 2 m (Fig.S1).

The soils at the study sites belong to the Dalhousie series, with the prevailing texture being a clay loam in the GEF experimental area during the 1997 and 1999 seasons and a silty clay loam in the CFIA experimental area during the other seasons (Table S1). The soil texture and chemical properties from experiments were measured as described by Crépeau et al. (2021). The soil water holding characteristics, including the lower limit of plant extractable soil moisture (LL), drained upper limit (DUL), and saturated soil water content (SAT) for each soil layer, were computed using pedotransfer functions for soils in temperate environments (Saxton et al., 1986; Gijsman et al., 2002). The SWC was measured using time domain reflectometry rods (TDR) installed at different depths (Table 1) (see Crépeau et al. 2021 for more details).

A total of six experiments were conducted during five soybean growing seasons, two of which took place in 2016, with early and late seeding dates (Table 1). The soybeans were grown under rainfed conditions and in rotation with maize, spring wheat, or spring canola, depending on the year. The fields were tile-drained, flat, and larger than 20 ha each. Maize and spring canola residues were incorporated, while spring wheat straw was baled. Tillage was done in the spring prior to seeding, using a disc harrow in 1997, a disc ripper in 2008, and a multi-finisher the remaining years, except for 1997 when soybean was seeded with no-tillage. Soybeans were sown at a seeding rate from 26.4 to 61.8 plants m^{-2} and with a 0.50 m row spacing from May 10 to June 3, depending on the year and treatment (Table 1). Soybean cultivars changed from year to year to cultivars of similar maturity (Table 1). AGB was measured by sampling 0.20–0.40 m^2 (15–20 plants) per site every 2 to 3 weeks, and samples were separated into leaf, stem, and pod tissue (Crépeau et al. 2021). LAI was measured destructively (LI3100; Li-Cor Inc., Lincoln, NE) or indirectly (LI2000 and LI2200; Li-Cor Inc., Lincoln, NE). Soybean yield was measured manually in 1997 and 1999 by harvesting 20 plants (about 0.4 m^2). In 2008 and onward, a combine harvester equipped with a yield monitor and moisture meter (9410, John Deere) was used to harvest the entire field.

2.2. Participating crop models

Participation in this study was based on the voluntary contribution of individual modeling teams who were experienced in using a particular process-based eco-physiological crop model. A total of six families of crop models with the ability to simulate soybean growth and development participated in this study (Table 2). The APSIM and DSSAT-CSM-CROPGRO (hereinafter CROPGRO) modeling teams provided additional simulations based on different methods of computing ET. Thus, a total of 15 flavors of models (based on six crop models with two flavors of APSIM and nine flavors of CROPGRO) were compared with each other and against measured data (Table 1).

Crop models participating in this study employed different approaches for estimating ET (see detailed description in Supplementary Information S1 and Table S2). Most models estimated ET by first calculating a potential crop evaporative demand (ETp) based on environmental variables and Priestley-Taylor (Priestley and Taylor, 1972) or FAO-56 Penman-Monteith (Allen et al., 1998) approaches. The ETp was calculated directly for soybean in some modeling scenarios, while others based ETp calculations on prior calculations of ET for a grass or alfalfa reference crop (ETo and ETr, respectively). Some models calculated potential crop transpiration (Tp) as a function of leaf area index (LAI) and an extinction coefficient, with the remaining energy being absorbed by the soil surface and contributing to potential soil water evaporation (Ep). Another approach employed the FAO-56 dual crop coefficient method (Allen et al., 1998) to calculate Ep and Tp from reference ET for the tall or short reference crop (CROPGRO ASCE models, Table 2). Overall, participating CROPGRO models in this study explored four distinct methods for simulating ETp (Table 2 and S2, Supplementary Information S1).

The ETp simulated by most models was used to determine the

Table 1
Description of experiments conducted at the Greenbelt Experimental Farm (GEF) or adjacent farm of the Canadian Food Inspection Agency (CFIA) Fallowfield Laboratory in Ottawa. The soil water contents available for each depth and experiment are reported (Yes/No).

Crop season	Cultivar and maturity group (MG)	Sowing date	Harvest date	Plant density (plants m^{-2})	Field	Depths of soil water content sensors (m)							
						0.06	0.15	0.25	0.30	0.50	0.70	0.90	1.10
1997	Maple Glen (0)	Jun 03	Oct 31	26.4	GEF	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
1999	2702R (0)	May 20	Oct 07	43.5	GEF	No	Yes	No	No	Yes	Yes	Yes	No
2008	P90M20 (0)	Jun 10	Oct 14	51.8	CFIA	Yes	Yes	Yes	No	Yes	Yes	No	Yes
2016 (Early sown)	P06T28R (0)	May 12	Nov 07	56.9	CFIA	Yes	Yes	Yes	No	Yes	No	No	No
2016 (Late sown)	P06T28R (0)	May 25	Nov 08	61.8	CFIA	Yes	Yes	Yes	No	Yes	Yes	No	Yes
2019	B150Y1 (1)	May 18	Nov 21	29.5	CFIA	Yes	Yes	Yes	No	Yes	No	No	No

Table 2
List of participating models and ET flavors with their acronyms. A detailed description of ET approaches used by models is provided in Table S1.

Acronym	Model Name and ET flavor	General model reference
APSIM	APSIM	Robertson and Carberry, 1998; Keating et al., 2003
APSIMNG	APSIM Next Generation	Holzworth et al., 2018
AQUACROP	AQUACROP	Steduto et al., 2009, Raes et al., 2009, Hsiao et al., 2009
CROPGRO-EBL	DSSAT-CSM-CROPGRO-Soybean Energy Balance	Cuadra et al., 2021
CROPGRO-FR	DSSAT-CSM-CROPGRO-Soybean FAO-56 Ritchie	Boote et al., 1998; Jones et al., 2003; Hoogenboom et al., 2019 ; DeJonge and Thorp, 2017; Ritchie et al., 1972; Suleiman and Ritchie, 2003
CROPGRO-FS	DSSAT-CSM-CROPGRO-Soybean FAO-56 Suleiman	
CROPGRO-RR	DSSAT-CSM-CROPGRO-Soybean Priestley-Taylor Ritchie	
CROPGRO-RS	DSSAT-CSM-CROPGRO-Soybean Priestley-Taylor Suleiman	
CROPGRO-SR	DSSAT-CSM-CROPGRO-Soybean ASCE-short crop Ritchie	
CROPGRO-SS	DSSAT-CSM-CROPGRO-Soybean ASCE-short crop Suleiman	
CROPGRO-TR	DSSAT-CSM-CROPGRO-Soybean ASCE-tall crop Ritchie	
CROPGRO-TS	DSSAT-CSM-CROPGRO-Soybean ASCE-tall crop Suleiman	
DNDC	DNDC	Zhang and Niu, 2016
LINTUL	SIMPLACE LINTUL	Wolf, 2012
MONICA	MONICA	Nendel et al., 2011

maximum threshold for daily ET under given environmental conditions, while the final simulated ET (ETs) and final simulated T (Ts) were dependent on soil water availability and the potential crop root water uptake. Some models did not rely on computing ETp to simulate ETs. One example is that of the CROPGRO-EBL model, which employed an energy balance approach that simulates energy and mass exchange in the crop's micro-climate to determine latent heat flux and associated water losses through evaporation (Cuadra et al., 2021). Another approach to simulate Ts and ETs was used by APSIM, which used simulated daily biomass growth and a transpiration coefficient approach to estimate Ts (Snow and Huth, 2004).

Models also employed different methods to simulate soil water evaporation (Es) from Ep (Supplementary Information S1, Table S2). Most models relied on the two-stage soil evaporation method by Ritchie et al. (1972) (referred to as "Ritchie", Table 2). Participating CROPGRO models in this study explored two methods for simulating soil evaporation: the "Ritchie" method and the method referred to as "Suleiman" (Table 2) based on upward flux calculations of soil moisture across soil layers (Suleiman and Ritchie, 2003, 2004).

In addition to different methods for simulating ETs, models participating in this study varied in levels of complexity for simulating soybean development and growth, types of vegetative and reproductive tissues simulated, and methods to account for environmental effects on in-season development, growth, and partitioning. A detailed description of model strategies for simulating soybean growth and development, and their response to environmental factors, can be found in Kothari et al. (2022, 2024).

2.4. Model calibration

The model calibration procedures were conducted in two stages: (i) step 1, calibration with crop growth and yield data, and (ii) step 2, calibration with measured ET and soil water content data. For step 1, the modelers were provided: (i) weather data, (ii) measured and/or estimated soil properties, (iii) timing of crop phenological stages, (iv) crop growth and yield data, (v) management information, and (vi) initial SWC. Additionally, modelers were allowed to calibrate soil parameters in step 1 as well. For step 2, daily ET and SWC were also provided, and modelers were allowed to calibrate crop growth coefficients if needed. This approach aligns with methods adopted in other AgMIP studies, which focus on assessing model performance under different levels of measured data for calibration, rather than separating datasets for independent calibration and evaluation (Rosenzweig et al., 2013; Kimball et al., 2019; Kothari et al., 2022, 2024; Wang et al., 2024). Our calibration approach has practical applications to quantify model uncertainty for simulation of daily ET when only crop growth data are available for calibration (most likely scenario for models), and the potential to improve simulation of daily ET after detailed calibration with soil moisture and daily ET data (see modified parameters in Table S3).

During step 1, the objective was to assess uncertainty in ETs simulations when only measured data on crop growth and development were available for model calibration. In step 1, modelers optimized phenology, crop growth, and partitioning coefficients for each cultivar while evaluating AGB and LAI and ensured reasonable simulated seed yield and mass of pod, leaf, and stem. The phenology data were scarce, and the modelers were, therefore, instructed to calibrate assuming all years shared the same cultivar and to rely on crop growth data to calibrate phenology coefficients. The modelers were allowed to further modify coefficients for any given year and create different cultivars for different years if needed for better simulation of in-season growth data. The modelers were also allowed to adjust some soil parameters based on their pedotransfer functions. In step 2, the aim was to parametrize and improve models to accurately simulate ETs under the study's environmental conditions and quantify the improvement in ETs and SWC simulations when daily ETs and SWC data were provided for calibration. In step 2, the modelers optimized parameters associated with crop water use, root water uptake, and soil water dynamics. The modelers could also modify the parametrization of crop growth coefficients, if necessary, during step 2 (see modified parameters between step 1 and step 2 in Table S3).

2.5. Data analysis and methods for evaluating model performance

After calibration step 1 and 2, the performance of individual models was evaluated for the simulation of daily ETs, cumulative ETs, LAI, AGB, leaf, stem, and pod weight, seed yield, and daily SWC as compared to measured data. In addition, the performance of a model ensemble calculated as the median of simulations across models (referred to as Median hereinafter) was evaluated. The median was used instead of the mean because the median is a more appropriate measure of central tendency, and it is not affected by extreme values. The purpose of evaluating the performance of the median across models was to quantify the applicability of a model ensemble to reduce uncertainty in ETs simulations. For this purpose, only two out of the nine CROPGRO flavors were included in the ensemble to avoid over-representation of CROPGRO model variation as compared to the variation found among groups of models. The two CROPGRO flavors included for computing the median across models were: (i) CROPGRO-RR because it is currently the default option, and (ii) CROPGRO-EBL because it uses a very different approach to calculating ETs.

Model performance was evaluated against measured data for the simulation of daily ETs, cumulative ETs, LAI, AGB, seed yield, and daily SWC using (i) the root mean square error (RMSE), (ii) normalized RMSE (nRMSE, Loague and Green, 1991), (iii) normalized bias (nBias), and (iv) index of agreement (D-statistic) (Willmott et al., 1985).

$$\mathrm{RMSE_m} = \sqrt{\frac{\sum_{i=1}^{N_s}\left(Y_i - \widehat{Y}_{m,i}\right)^2}{N}} \tag{1}$$

$$\mathrm{nRMSE_m} = \sqrt{\frac{\sum_{i=1}^{N_s}\left(Y_i - \widehat{Y}_{m,i}\right)^2}{N}} \times \frac{100}{\overline{Y}} \tag{2}$$

$$nBias_m = 100\frac{Y_i - \widehat{Y}_{m,i}}{Y_{m,i}} \tag{3}$$

$$D_m = 1 - \frac{\sum_{i=1}^{N}\left(Y_i - \widehat{Y}_{m,i}\right)^2}{\sum_{i=1}^{N_s}\left(\left|\widehat{Y}_{m,i} - \overline{Y}\right| + \left|Y_i - \overline{Y}\right|\right)^2} \tag{4}$$

where Y_i is the ith measured value and $\hat{Y}_{m,i}$ is the corresponding simulated value by model m; $\bar{Y}$ is the average of measured values; and N is the number of measured values.

The nBias quantifies the variation between simulated and measured values (%), if the model under-estimates, the value for nBias is negative, while if the model over-estimates, the value for nBias is positive. The D-statistic is the degree to which the simulated model values are approached by the measured values; it ranges from 0 to 1, with 0 indicating no agreement between the measured and simulated values and 1 indicating perfect agreement. A high value for the D-statistic, a low value for RMSE, nRMSE, and nBias near zero indicate better model performance.

Before computing statistics for model evaluation, the SWC data were grouped into three soil depth intervals (Table S2): (i) 0.05 to 0.25 m, (ii) 0.25 to 0.50 m, and (iii) 0.50 to 1.00 m. For total crop growing season ET, the same period was considered across all models, from the planting date to the measured harvest date.

To further evaluate model performance for simulation of daily ETs, statistics were computed after separating daily ETs data into four different periods: (i) before emergence (20 days before emergence to emergence); (ii) incomplete canopy (from emergence to the day when LAI reached a value of 2.5); (iii) full canopy (from the day when LAI was greater than 2.5 until 30 days later); and (iv) late reproductive (end of full canopy phase to R8). The mean square error (MSE) was broken down into three additive components: standard bias (SB), lack of correlation (LC), and non-unity of slope (NU) (Gauch et al., 2003).

$$\mathrm{MSE_{m,p}} = \frac{\sum_{i=1}^{N_s}\left(\widehat{Y}_{m,p,i} - Y_{p,i}\right)^2}{N_p} \tag{5}$$

$$\mathrm{LC_{m,p}} = \left(1 - r^2\right)\frac{\sum_{i=1}^{N_p}\left(Y_{p,i} - \overline{Y}\right)^2}{N_p} \tag{6}$$

$$\mathrm{NU_{m,p}} = (1 - b)^2\frac{\sum_{i=1}^{N_p}\left(\widehat{Y}_{m,p,i} - \overline{\widehat{Y}_{m,p}}\right)^2}{N_p} \tag{7}$$

$$\mathrm{SB_{m,p}} = \left(\frac{\sum_{i=1}^{N_p}\left(\widehat{Y}_{m,p,i}\right)}{N_p} - \frac{\sum_{i=1}^{N_p}\left(Y_{p,i}\right)}{N_p}\right)^2 \tag{8}$$

where the MSE is specified for the model m during period p. The $Y_{p,i}$ is the ith measured value during period p, and $\hat{Y}_{m,p,i}$ is the corresponding simulated value for model m; $\bar{Y}$ is the average of measured values during period p; $\overline{\widehat{Y}_{m,p}}$ is the average of the simulated value for model m; and N_p is the number of measured values at the site.

We also evaluated the relationship between the performance of models for simulating daily ET and their performance simulating aboveground biomass (AGB), and LAI using regression analysis. Pearson correlation coefficients were obtained from the relationship between the nRMSE for simulation of daily ET, and the nRMSE for either AGB or LAI, with statistical significance considered at p-values of 0.1 or lower (the lowest observed p-value was 0.08) (Cohen et al., 2009). The nRMSE for the simulation of daily ET, LAI, and ABG computed by model and growing season were employed for this analysis.

We evaluated the performance of different soil evaporation and potential ET methods within the CROPGRO flavors by analyzing the bias and the nRMSE for the simulation of daily ET with an analysis of variance (ANOVA) using proc mixed in SAS v 9.4 (SAS Institute, Cary, North Carolina, USA). The daily bias and the nRMSE computed by growing season were utilized for this analysis. The model included as fixed factors the soil evaporation method (Ritchie or Suleiman), the potential ET method (Priestley-Taylor, FAO-56, ASCE-Short crop, and ASCE-Tall crop), the canopy developmental stage (before emergence, incomplete canopy, full canopy, and late reproductive), and their interaction. Growing season was included as random factor in the model for the analysis of nRMSE in addition to day of year nested within growing season for the analysis of the bias. This analysis included only results after calibration step 2.. The CROPGRO-EBL model was excluded from the analysis since it has a substantially different approach to simulating ET that did not share any soil evaporation or potential ET method with the other approaches, making our analysis unbalanced. For significant fixed factors at $p < 0.05$, differences between means were explored with least significant differences using the LSMEANS statement.

3. Results and discussion

3.1. Performance of crop models for simulating daily and cumulative ETs

After calibration step 1, the simulated daily ETs averaged 2.60 mm d^{-1}, and daily values ranged from 0.30 to 8.20 mm d^{-1} depending on the model; the measured values averaged 1.67 mm d^{-1} and daily values ranged from 0.01 to 5.48 mm d^{-1} across growing seasons. The visual comparison of the median showed that the values of the daily time-course simulations were consistently larger than the ET values measured using eddy covariance (Fig. 1). Nevertheless, the median was close to measurements until about 50 days after sowing for all growing seasons.

After calibration step 2, the time course of simulated daily ETs was still consistently larger than the measured ET values using eddy covariance for all growing seasons (Fig. 1). The ETs averaged 2.66 mm d^{-1} after calibration step 2, with a range of 0.30 to 8.25 mm d^{-1} for the daily values among the models (Fig. 1). However, changes in crop water use and soil water parameters in some models during step 2 resulted in a relative decrease in the ETs of the ensemble median compared with the ensemble median ETs during step 1.

When evaluating the performance of individual models simulating daily ETs, the RMSE ranged from 0.8 to 2.6 mm d^{-1} among models with an average RMSE of 1.6 mm d^{-1} after calibration step 1 (Fig. 2a). After step 2, the RMSE ranged from 0.7 to 2.8 mm d^{-1} depending on the model, with an average of 1.4 mm d^{-1} (Fig. 2a). The top five performing models with lowest RMSE for simulation of daily ETs after step 2 were DNDC, the ensemble median, MONICA, APSIM, and APSIMNG. The D-statistic for the simulation of daily ETs across models ranged from 0.5 to 0.9 with an average of 0.7 for both steps. The DNDC had the best performance for the simulation of daily ETs, with a D-Statistic of 0.9, followed by a group of models of similar performance (ensemble median, MONICA, APSIM, and APSIMNG), with a D-statistic value of 0.8 (Fig. 2b).

The measured total ET during the crop growing season (from emergence until R8) averaged 250 mm and ranged from 179 to 330 mm depending on the season. The simulated growing season ETs during the same period averaged 372 mm and ranged from 216 to 615 mm. Thus, the models overestimated the growing season ETs on average. The RMSE for simulation of growing season ETs ranged from 75 to 242 mm, with an average RMSE of 150 mm after calibration step 1; and ranged from 40 to 289 mm, with an average RMSE of 123.4 mm for step 2 (Fig. 2c).

When evaluating the effect of calibration across models, the APSIM and MONICA models had a notable improvement in the daily and

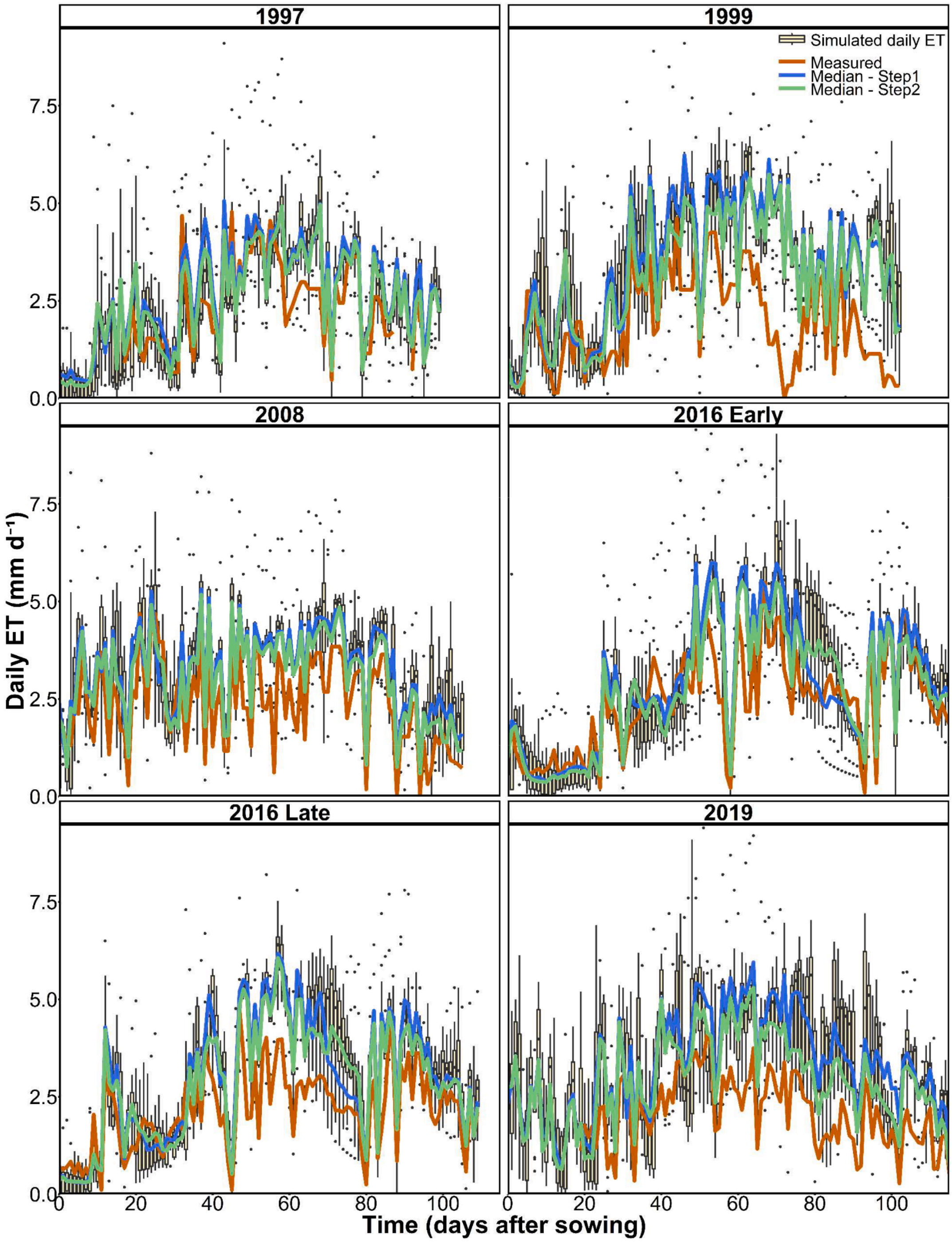

Fig. 1. Boxplots of simulated daily evapotranspiration (ETs) across all models after calibration step 2 during different soybean growing seasons in Ottawa, Canada. The measured daily ETs (orange line) was estimated based on the eddy covariance method. Median ETs values for step 1 (blue line) and step 2 (green line) were computed from APSIM, APSIMNG, AQUACROP, CROPGRO-RR, CROPGRO-EBL, LINTUL, DNDC, and MONICA models. Boxplots are set at 90th (upper whisker), 75th (upper quartile), 50th (median), 25th (lower quartile), and 10th (lower whisker) percentiles; outliers are shown as black dots.

growing season ETs simulation after calibration step 2. In these models, the RMSE decreased by 58% for daily ET and 65% for growing season ETs between calibration steps (Fig. 2a and 2c). As discussed later in this manuscript, the improvement in the simulation of ETs after calibration was related to changes in model parametrization that affect crop water use, but also changes in soil water parameters or re-calibration of crop growth coefficients.

The DNDC model was the most accurate for the simulation of both daily and total crop-growing season ETs. This model was previously used to simulate ET for a humid continental climate in eastern Canada and it had been evaluated and improved for the simulation of ETs in corn and wheat with eddy-covariance ET flux data (Sansoulet et al., 2014; Dutta et al., 2016), and for soybean and other crops with a comparison with water budget measurements (Guest et al., 2018). In the first versions of the DNDC model, the method to estimate ETp was based on an adaptation of the Thornthwaite equation (Li et al., 1992), which uses only daylength and temperature as weather inputs. Sansoulet et al. (2014) found that early versions of DNDC with this approach to computing ETp exhibited a 15% overestimation in ETs compared with measured data. In the current version of the DNDC model (DNDC95 v.CAN), the FAO-56

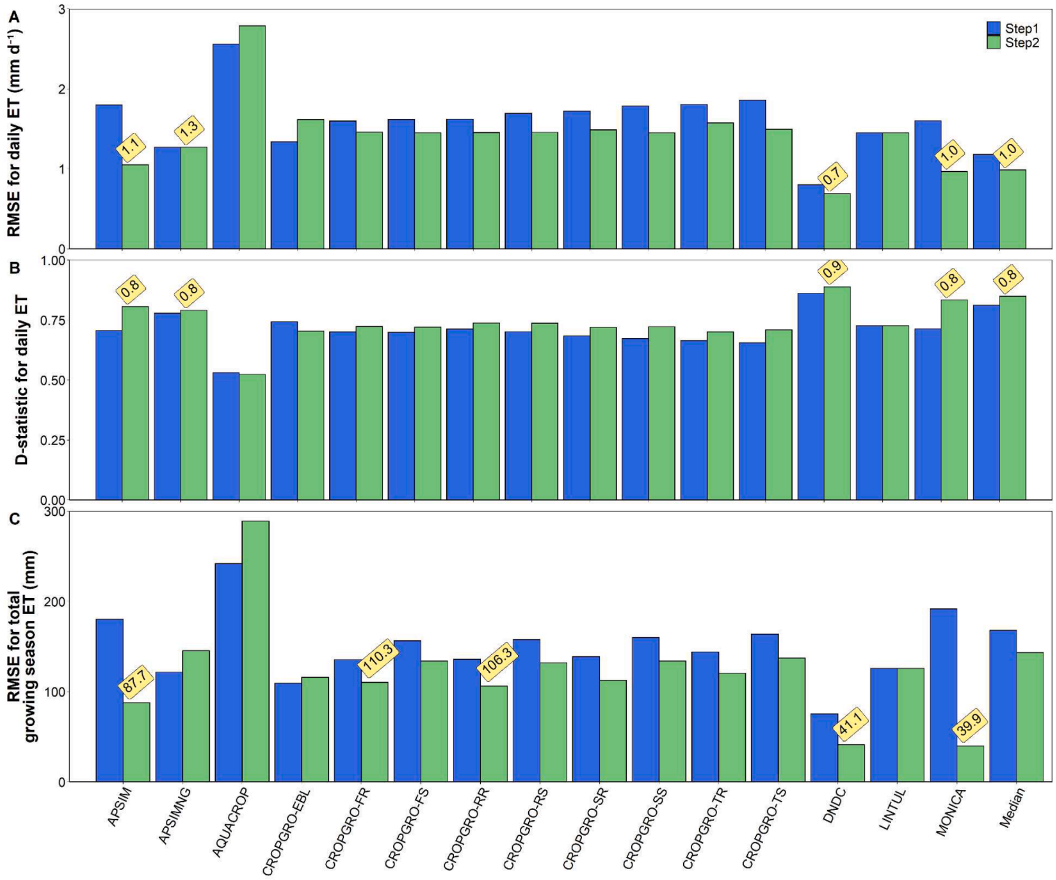

Fig. 2. A) Root mean square error (RMSE) for the simulation of daily evapotranspiration (ETs), B) D-statistic for daily ETs, and C) average RMSE for total growing season ETs. Statistics were obtained across all growing seasons (1997, 1999, 2008, 2016 early, 2016 late, and 2019). Median values were computed from eight of the models (APSIM, APSIMNG, AQUACROP, CROPGRO-RR, CROPGRO-EBL, LINTUL, DNDC, and MONICA). The top five performing models, after step 2, for RMSE and D-statistic are shown with yellow flags.

Penman-Monteith method was implemented to simulate ETp as a function of ETo and crop coefficients (Kc), which reduced ETs estimates for annual crops (Dutta et al., 2016). Furthermore, the model was later improved for the simulation of crop water uptake by increasing maximum soil depth from 0.50 to 2.00 m, including root penetration and density functions, incorporating a fluctuating water table, and option for tile drainage (Smith et al., 2020). The fact that the DNDC model was developed and improved under cool weather conditions considering water movement in frozen soils likely contributed to its best performance in this study (Dutta et al., 2018). MONICA obtained the best performance for cumulative ETs after step 2. Between step 1 and step 2, the RMSE for the simulation of daily ETs in MONICA dropped from 191.5 to 39.9 mm, partially due to the adjustment of Kc, which directly affected the simulation of potential ET. The crop-specific maximum root depth parameter was also adjusted in the MONICA model during Step 2, which was shown to have high sensitivity for the simulation of crop growth in previous studies (Battisti et al., 2017).

One of the objectives of the study was to assess the effectiveness of a model ensemble to reduce the uncertainty in the simulation of daily ETs. Based on our findings under a warm-summer humid continental climate, the ensemble median achieved the second-best performance for the simulation of daily ETs, ranking after the DNDC model after both calibration steps. After calibration step 2, the ensemble median was tied in second place with the MONICA model (Fig. 2a). Thus, the ensemble median was useful in reducing high uncertainty associated with individual models for simulation of ETs after step 1, when only crop growth data were available for calibration. After calibration step 2, other models improved performance for the simulation of daily ETs, but the ensemble was still useful in reducing high uncertainty. Of interest, the ranking of the ensemble median for the simulation of total growing season ETs was worse (13 and 14th for steps 1 and 2, respectively) than for the ranking of the ensemble median for the simulation of daily ETs. This could be attributed to the potential compounding effect of errors in the individual models over time, as errors that oscillate in direction throughout the growing season may offset each other when calculating daily ETs but not when calculating total growing season ETs. The ensemble median consistently underestimated ETs, leading to a cumulative amplification of errors over time and worse performance in simulating total ETs despite its good performance in simulating daily ETs. Overall, the ensemble was useful in reducing uncertainty for the simulation of ETs, but it did not rank first. These results contrast with those from a multi-model evaluation of maize models by Kimball et al. (2019), who reported the best performance by the ensemble median across different calibration steps in their study. Their superior performance of the median may be partly due to the greater number of models used (29 maize models) in the study by Kimball et al. (2019) compared to our study (the ensemble median was calculated from eight soybean models). Even when computing the ensemble median for all 15 flavors of soybean models, including all flavors of the CROPGRO model, the ensemble median failed to rank first in our study (data not shown). Thus, a major factor influencing the performance of a model ensemble in our study is the fact that most models simulated greater ETs compared to measured values, so the simulated ETs were not distributed around the measured but had a consistent bias (Fig. 1).

3.2. Model performance for simulation of daily ETs during different periods

Model performance was analyzed for the simulation of daily ETs during different periods to identify potential sources of model uncertainty (Fig. 3a). For instance, during the 20 days prior to crop emergence, model uncertainty in the simulation of ETs was attributed solely to a model error in the simulation of Es. Measured daily Es during the period before crop emergence averaged 1.3 mm d^{-1}, and simulated daily Es across models averaged 1.6 and 1.4 mm d^{-1} for steps 1 and 2, respectively. During the period of incomplete canopy, when the ETs rate is a combination of Es and Ts, measured ET averaged 2.1 mm d^{-1}, and simulated ETs averaged 3.2 mm d^{-1} and 2.9 mm d^{-1} for steps 1 and 2, respectively. During the period of full canopy, most of the ETs consisted of Ts, and measured ET averaged 2.5 mm d^{-1}, while simulated ETs during the full canopy phase averaged 4.1 and 3.8 mm d^{-1} for steps 1 and 2, respectively. Thus, the difference between measured and simulated ETs was greater during the full canopy phase. During the late reproductive phase, when LAI was declining due to rapid senescence, measured ET decreased to 2.1 mm d^{-1} on average, and simulated ETs averaged 3.0 mm d^{-1} and 2.8 mm d^{-1} for steps 1 and 2, respectively (Fig. 3a).

Overall, the uncertainty in daily ETs simulations decreased with calibration during the four periods. The average RMSE for simulation of daily ETs after step 2 was consistently lower than after step 1 for all the models and periods that were evaluated (Fig. 3b). The average RMSE for simulation of daily ETs ranged from 1.10 to 2.03 mm d^{-1} among models and different periods for step 1, and from 1.00 to 1.76 mm d^{-1} for step 2 (Fig. 3b). In general, when the contribution of Ts towards total ETs was greater than that of Es, the RMSE for simulation of ETs was greatest. These results are expected due to daily ET being the least during the period prior to emergence when there is only E.

When evaluating model performance for the simulation of daily ETs by period with the nRMSE (Fig. 4), the results revealed a greater uncertainty for the simulation of Es during the period before emergence (nRMSE averaged 84.6% for step 1, and 76.8% for step 2), compared to the simulation of ETs during other periods (nRMSE averaged 76.6% for step 1 and 66.3% for step 2 across incomplete and full canopy periods). In general, step 2 of the calibration reduced the nRMSE for daily ETs during most periods and for most models. Two exceptions were the AQUACROP and the CROPGRO-EBL models, which showed an increase in nRMSE for daily ETs in all periods after calibration step 2 (Fig. 4). Within the CROPGRO models, there was also a relative increase in the nRMSE for daily ETs after step 2 during the late reproductive phase, for those model flavors employing the "Suleiman" method (Suleiman and Ritchie, 2003, 2004) for the simulation of Es (Fig. 4). In contrast, the nRMSE for daily ETs decreased after Step 2 for the CROPGRO models employing the "Ritchie" method (Ritchie et al., 2009) for calculating Es (Fig. 4). Overall, the DNDC model performed best for the simulation of daily ETs during incomplete canopy, full canopy, and late reproductive periods and after step 2, which can be attributed to the abovementioned ET and hydrologic processes in the model previously developed and tested against measured ET data under the same warm-summer humid continental climate (Sansoulet et al., 2014; Dutta et al., 2016).

The ranking of the model ensemble and its ability to reduce uncertainty simulating ETs were dependent on the period evaluated (Fig. 4). Overall, the ensemble median exhibited the best performance in simulating daily ETs during the period prior to emergence and ranked among the top five models in the remaining periods. After step 1, the ranking of the ensemble median was 1st during the period before emergence (nRMSE= 47%), followed by the DNDC model (nRMSE= 49%). During the period of incomplete canopy after step 1, the DNDC (nRMSE = 37%) and CROPGRO-EBL (nRMSE = 45%) ranked better than the ensemble median (nRMSE = 56 %). In the period of full canopy after step 1, the DNDC (nRMSE = 38 %) and LINTUL (nRMSE = 59%) models ranked better than the ensemble median (nRMSE= 37%) for simulation of daily ETs. During the late reproductive period, the model ensemble ranked 4th, with an nRMSE for daily ET of 58%, and outperformed by DNDC (nRMSE = 45%), LINTUL (nRMSE =54%), and APSIMNG (nRMSE =54%). The high ranking of the ensemble median for the period of soil water evaporation is consistent with the results of the maize multi-model study by Kimball et al. (2019). However, the performance of the ensemble median was worse for periods including transpiration from the soybean crop, unlike in the maize study by Kimball et al. (2019), where the ensemble median showed the best performance for the simulation of ET across different periods and calibration steps in the multi-model study with 29 maize models. Thus, the authors concluded that a multi-model approach would be best for simulating ET rather than using any single maize model (Kimball et al., 2019). Based on the results for soybean simulations under a humid continental climate, models with

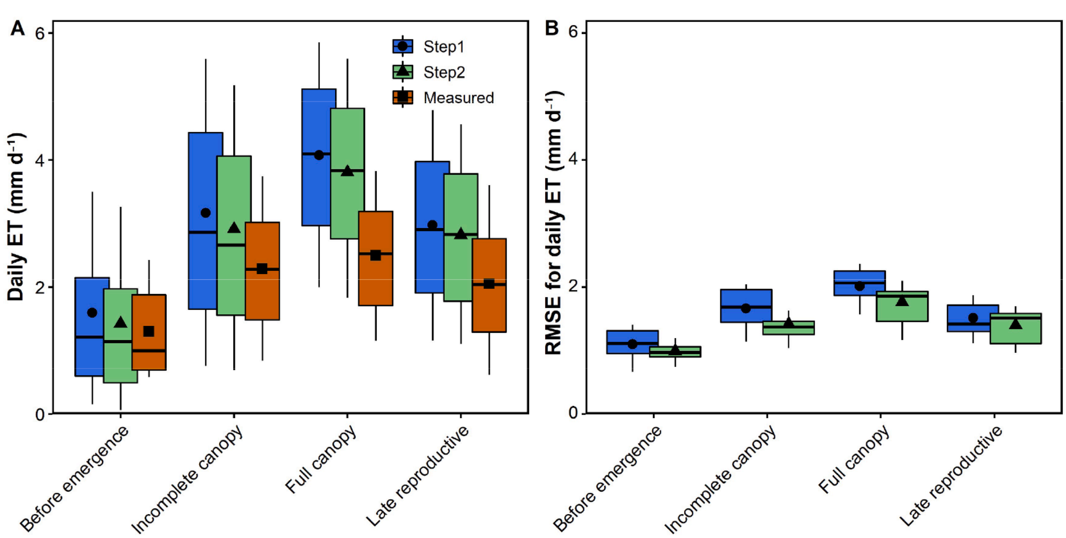

Fig. 3. Boxplots of daily evapotranspiration (ET) across different growing seasons simulated by models and measured by eddy covariance (A), and root mean square error (RMSE) between measured and simulated daily ET (B). Daily ET data were analyzed during four different periods: (i) before emergence (20 days before emergence to emergence), (ii) incomplete canopy (emergence to the day of LAI reach a value of 2.5), (iii) full canopy (day from LAI beyond 2.5 until 30 days later), and (iv) late reproductive (end of full canopy phase to R8). Boxplots are set at 90th (upper whisker), 75th (upper quartile), 50th (median), 25th (lower quartile), and 10th (lower whisker) percentiles, and the symbols (circle, triangle, or square) represent the average.

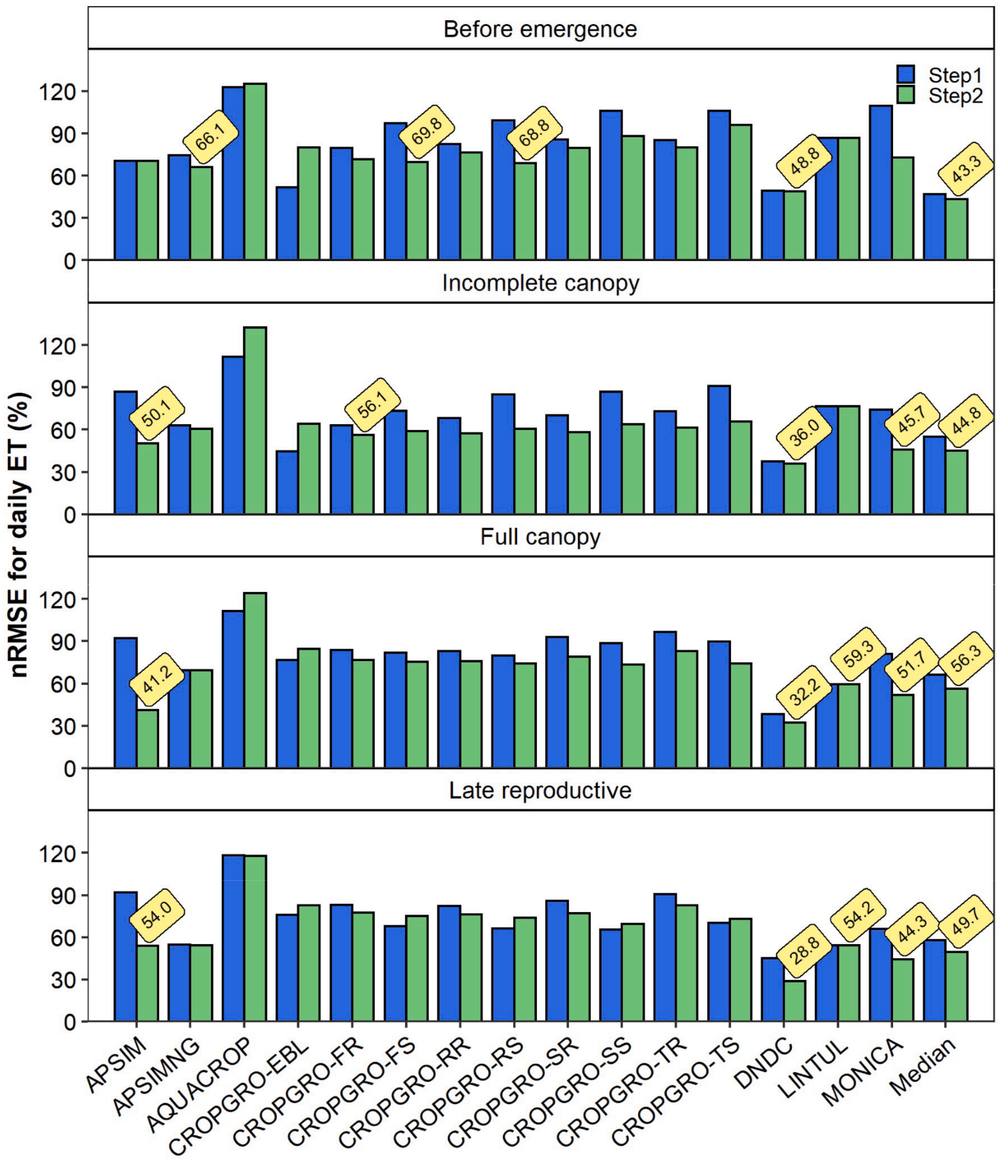

Fig. 4. Normalized root mean square error (nRMSE) for daily evapotranspiration (ETs) simulated by 15 soybean models (based on six crop models with two flavors of APSIM and nine flavors of CROPGRO) for periods: (i) before emergence (20 days before emergence to emergence), (ii) incomplete canopy (emergence to day of LAI reaching value of 2.5), (iii) full canopy (day of LAI beyond 2.5 until 30 days later), and (iv) late reproductive (end of full canopy phase to R8). Ensemble median values were computed from the models APSIM, APSIMNG, AQUACROP, CROPGRO-RR, CROPGRO-EBL, LINTUL, DNDC, and MONICA. The top five performing models, after step 2, for nRMSE are shown with yellow flags.

prior experience under a given climate may be best to reduce uncertainty in the simulation of soybean ET compared to a model ensemble.

The model performance was also evaluated for the simulation of daily ETs by period with the nBias (Fig. 5). The nBias for daily ETs revealed that for most models and periods, the daily ETs were higher on average compared to the measured ET (Fig. 5). However, model calibration in Step 2 reduced nBias for daily ETs compared to model simulations after Step 1, with few exceptions. During the period before emergence, the CROPGRO group of models using the Ritchie soil evaporation method had a lower nBias for daily ETs (−8.4%) compared to the Suleiman method (27.0%). In addition, the CROPGRO-EBL model had a negative nBias (−41.1%) in contrast with the positive nBias found with most other models during the period before emergence. DNDC showed the best performance for the simulation of daily ETs, with a nBias of 4.8% during the period before emergence.

During the period of the incomplete canopy, all models except DNDC overestimated daily ETs compared to the measured daily ET, with a nBias ranging from 9.2% to 44.2% depending on the model after step 1, and from 9.5% to 50.1% after step 2 (Fig. 5). In contrast, the DNDC model simulated daily ETs close to the measured on average after Step 1 (nBias = 2.8%), and with a relatively small underestimation after step 2 (nBias = −5.4%).

During the period of a full canopy, all models overestimated daily ETs on average relative to the measured daily ET, with a nBias ranging from 17.1% to 49.0% depending on the model after step 1. A group of three models (APSIM, DNDC, and MONICA) reduced nBias to values below 12% after Step 2, while the other models demonstrated a large overestimation of daily ETs compared to the measured data (nBias ranged from 28.1% to 52.1% depending on the model).

During the late reproductive phase, most models continued to simulate daily ETs values greater than the measured data, with an nBias ranging from 22.7% to 49.3% after Step 1 (Fig. 5). One exception was the APSIMNG model that underestimated daily ET after Step 1 (nBias = −13.6%). After step 2, the APSIM and the DNDC models simulated daily

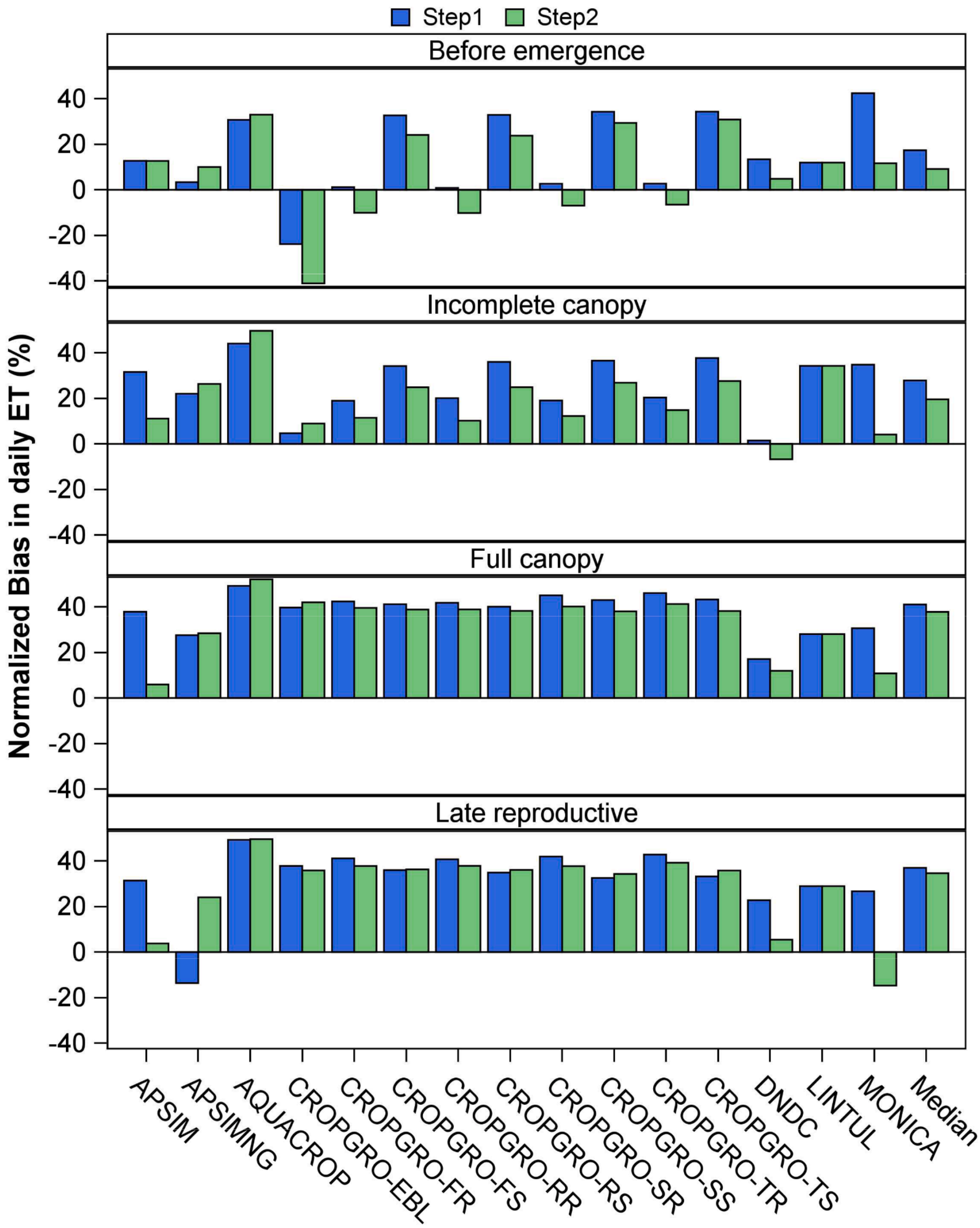

Fig. 5. Normalized bias (nBias) for the simulation of daily evapotranspiration (ETs) by 15 flavors of soybean models (based on six crop models with two flavors of APSIM and nine flavors of CROPGRO) for periods: (i) before emergence (20 days before emergence to emergence), (ii) incomplete canopy (emergence to day of LAI reaching value of 2.5), (iii) full canopy (day of LAI beyond 2.5 until 30 days later), and (iv) late reproductive (end of full canopy phase to R8). Ensemble median values are computed from the models APSIM, APSIMNG, AQUACROP, CROPGRO-RR, CROPGRO-EBL, LINTUL, DNDC, and MONICA.

ETs close to the measured data (nBias = 3.8% and 5.4%, respectively), while the MONICA model underpredicted daily ET on average (nBias = −14.7%). The remaining models continued to show a large positive nBias for the simulation of daily ET after calibration step 2 (nBias ranged from 24.0% to 49.6%).

Overall, the results are consistent with a prior evaluation of the STICS model for the simulation of ET with the same eddy flux dataset, where the crop coefficient method with Penman (1948) considerably over-estimated relative to measured ET (Crepeau et al., 2021). A resistance approach to calculate ETs based on Shuttleworth and Wallace (1985) improved simulation of daily ETs (nRMSE=36%) and reduced nBias (15%) compared to computing ETo based on Penman (1948) combined with a single surface crop coefficient approach (Maximum crop coefficient was 1.3) to compute ETp (nRMSE=67%; 34%), however, a bias between simulated ETs and measured ET persisted with both ET approaches after model calibration (Crepeau et al., 2021).

3.3. Analysis of different CROPGRO ET approaches

The analysis of variance of the bias and the nRMSE for simulation of daily ET within CROPGRO subtypes revealed no interaction between the soil evaporation and the potential ET method on the statistics evaluated (Table S4). Thus, this analysis was useful to study main effects of the soil evaporation and potential ET methods across the eight CROPGRO subtypes (excluding CROPGRO-EBL). The effect of the soil evaporation method on both the bias and the nRMSE was dependent on the duration of complete canopy coverage (Table S4). The Ritchie method had a lower bias and nRMSE compared to the Suleiman method, with the difference between the two methods being more pronounced during the period prior to emergence and the incomplete canopy period (Table 3). Before emergence, the Ritchie method achieved a near-zero bias (−0.04 mm) and a lower nRMSE (83.3%) compared to the Suleiman method (bias = 0.55 mm; nRMSE = 95.3%). During the incomplete canopy phase, when both soil water evaporation and transpiration contribute to ET, the Ritchie method again outperformed the Suleiman method, with lower bias (0.42 mm vs. 0.83 mm) and nRMSE (60.9% vs. 70.9%). In the full canopy phase, where transpiration dominates, the Ritchie method showed a smaller bias (1.53 mm vs. 1.64 mm) and a marginally lower nRMSE (79.1% vs. 83.5%). During the late reproductive phase, when LAI declines due to senescence, the Ritchie method also had lower bias (1.04 mm vs. 1.17 mm) and nRMSE (67.8% vs. 81.5%) compared to the Suleiman method. These results demonstrate that the Ritchie method provided more accurate simulations of daily ET across all crop development periods, likely due to its improved representation of soil evaporation dynamics under varying canopy conditions.

When evaluating potential ET methods within the CROPGRO flavors, there was a main effect of the potential ET method on the daily bias (Table S4). While the Priestley-Taylor, FAO-56, and ASCE-Short crop methods performed similarly, the ASCE-Tall crop method exhibited a relatively higher bias (mean bias = 0.94 mm) compared to other methods (bias ranging from 0.86 to 0.89 mm; Table 4). While the nRMSE was also higher on average for the ASCE-Tall crop compared to other methods, this difference was not significant (Table 4). These findings indicate that, under the same crop and soil parametrization, the choice of potential ET method within CROPGRO flavors had a limited impact on the accuracy of daily ET simulations. Thus, all potential ET approaches within CROPGRO can be recommended for soybean daily ET simulations based on our results. Within ASCE approaches, the ASCE-Short crop was best suited to simulate daily ET in a soybean canopy, reducing the systematic bias compared to the ASCE-Tall crop method.

3.4. Analysis of error decomposition for simulation of daily ETs by period

Partitioning model error for simulation of daily ETs into different types of error and by time period provided insights on sources of model uncertainty for simulating daily ET. The MSE for the simulation of daily ETs was partitioned into lack of correlation (LC), non-unity of slope (NU), and standard bias (SB) in Fig. 6. During the period before emergence and incomplete canopy, NU was the main component of the MSE for both steps, when ET was driven predominantly by E. Most of the models were either over-estimating or under-estimating daily ET values, but the degree of over- or under-estimation was not consistent across the range of measured values. In contrast, during the full canopy and late reproductive periods, the main component of the MSE was SB, indicating that most of the models had a consistent bias in their simulated daily ETs values relative to the measured data.

In general, for individual models and the ensemble median, a major component of the MSE was SB in step 1 and step 2 (Fig. 6). Accordingly,

Table 3
Mean bias and nRMSE for the simulation of daily ET by canopy period and soil evaporation method within CROPGRO flavors. Different letters within a canopy period indicate different means between soil evaporation methods at $p < 0.05$.

Canopy Period	Soil Evaporation method	Mean bias (mm)	nRMSE (%)
Before emergence	Suleiman	0.55 a	95.3 a
Before emergence	Ritchie	−0.04 b	83.3 b
Incomplete canopy	Suleiman	0.83 a	70.9 a
Incomplete canopy	Ritchie	0.42 b	60.9 b
Full canopy	Suleiman	1.64 a	83.5 a
Full canopy	Ritchie	1.53 b	79.1 a
Late reproductive	Suleiman	1.17 a	81.5 a
Late reproductive	Ritchie	1.04 b	67.8 b

Table 4
Mean bias and nRMSE for the simulation of daily ET by potential ET method within DSSAT-CROPGRO models. Different letters indicate different means between potential ET methods at $p < 0.05$.

Potential ET method	Mean Bias (mm)	nRMSE (%)
Priestley-Taylor	0.86 b	76.4 a
FAO-56	0.88 b	75.2 a
ASCE-Short	0.89 b	78.4 a
ASCE-Tall	0.94 a	81.1 a

the results indicate that the major source of error for the simulation of daily ETs was due to the magnitude of the simulated daily ET values, with a consistent bias (error attributed to SB) during periods when ET was mostly partitioned to T. Overall, the results indicate a strong tendency of most models to overestimate daily ETs in relation to data obtained by the eddy covariance method during the periods when ET is mostly composed of T.

A relatively large portion of the MSE for the simulation of daily ETs was associated with NU in some models, particularly during the period before emergence and incomplete canopy cover. This indicates that models may tend to overpredict ET more during days with relatively low or high ET demand. The data were further explored using individual regression analyses by model, comparing simulated and measured daily ET (Fig. S2-S5). The regression analysis revealed that most models simulated daily ETs values below the 1:1 line of equality, meaning they consistently overestimated daily ET compared to values measured using eddy covariance (Fig. S2-S5).

During the period before emergence, models with high NU (Fig. 6) overestimated ET on days when measured ET was relatively high (Fig. S2). Similarly, in other periods, most models overestimated ET to a greater extent on days with greater measured daily ET (Fig. S3-S5). For the MONICA model, simulated ETs values tended to be smaller than measured ET under conditions of relatively low daily ET, and overestimated under conditions of relatively high ET. For the DNDC model, the regression of simulated ETs against measured ET had a slope close to the 1:1 line (Fig. S2-S5), consistent with most of the MSE attributed to LC for this model.

Overall, the results indicate that the large MSE, SB, and NU in most models for the simulation of ET were due to a strong tendency to overestimate daily ET compared to measured data obtained by the eddy covariance method, particularly on days with relatively greater average ET. The results also reveal that model overpredictions were relatively greater during periods when ET was predominantly partitioned into crop transpiration. The bias between simulated ETs and measured ET is consistent with results from a prior evaluation of the STICS model at the same location for simulation of daily ET for a soybean crop (nBias=15–34% depending on the ET method; Crepeau et al., 2021). In contrast, the evaluation of the same ET methods within the STICS model for simulation of ETs in maize at the same location revealed a relatively smaller nBias (6–26%, averaged by ET method across wet growing seasons when LAI>1) (Saadi et al., 2022).

3.5. Differences across models simulating daily ETp and Tp

The model uncertainty in the simulation of daily ETs may be attributed to uncertainty in the computation of ETp and Tp, or to how models simulate soil water dynamics, root water uptake, and the associated ability to meet Tp demand and simulate final ETs and Ts (Figueiredo Moura da Silva et al., 2022; Chatterjee and Anapalli, 2023; Dias et al., 2023). Thus, model outputs after calibration Step 2 were compared for the simulation of daily ETp and Tp demand, as well as final ETs and Ts (Fig. 7).

The simulated daily ETp varied greatly across models and periods. For instance, the average daily ETp ranged from 1.6 to 6.4 mm d^{-1} depending on the model during the period before emergence, and from

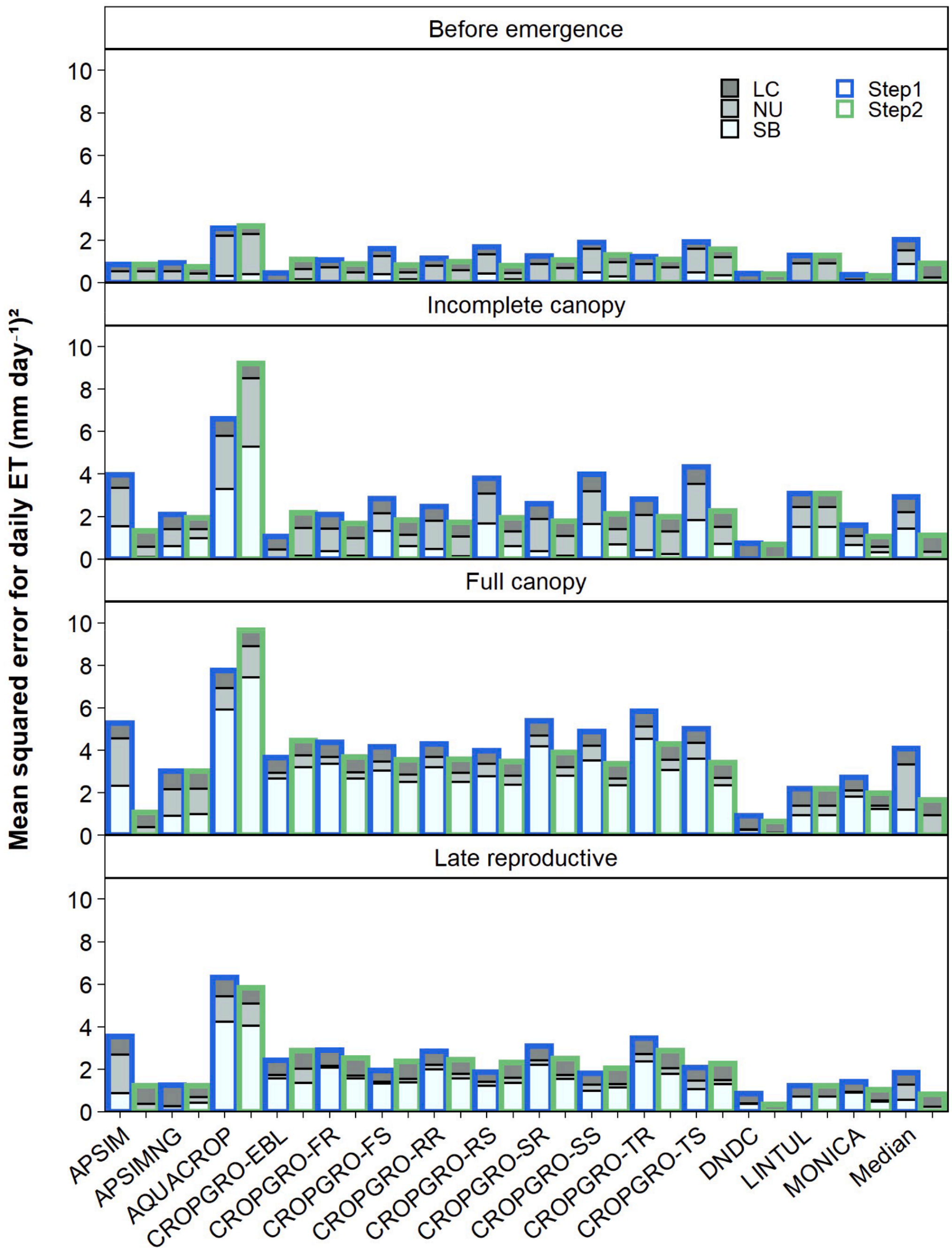

Fig. 6. Mean squared error (MSE) between simulated and mesaured daily evapotranspiration (ET). The bars show the components of mean squared error: lack of correlation (LC), non-unity of slope (NU), and standard bias (SB). Values reflect 15 flavors of soybean models (based on six crop models with two flavors of APSIM and nine flavors of CROPGRO) for periods: (i) before emergence (20 days before emergence to emergence), (ii) incomplete canopy (emergence to the day of LAI reaching value of 2.5), (iii) full canopy (day of LAI beyond 2.5 until 30 days later), and (iv) late reproductive (end of full canopy phase to R8). Ensemble median values were computed from the models APSIM, APSIMNG, AQUACROP, CROPGRO-RR, CROPGRO-EBL, LINTUL, DNDC, and MONICA.

3.1 to 6.6 mm d^{-1} during the period of full canopy. The daily ETp was the lowest during the late reproductive period, ranging from 2.3 to 5.3 mm d^{-1} depending on the model.

The simulated daily Tp was lowest during the period of incomplete canopy cover, ranging from 0.8 to 2.8 mm d^{-1} on average, depending on the model, and daily Tp was highest during the period of full canopy, ranging from 1.7 to 4.7 mm d^{-1}. Of interest, the daily Ts was very close to Tp, indicating that the soybean crop experienced very limited drought stress during model simulations. Thus, differences in the simulation of daily Tp were driving differences in simulated daily Ts and ETs (rather than simulation of soil water dynamics and root water uptake).

The results evidenced pronounced differences in the simulation of daily ETp across models that are consistent with the results from Battisti et al. (2018) who compareded a group of four soybean models (APSIM, AQUACROP, DSSAT-CROPGRO, MONICA). Model intercomparisons for the simulation of ET in maize also revealed large differences in the simulation of ETp across models, which was an important driver for the high variability in simulated ET (Kimball et al., 2019, 2023). When evaluating the different "upstream" variables used in the calculation of daily ETp reported by models, Kimball et al. (2023) found that 13 out of 44 models reported ETo and that it varied by a factor of 2 across models during mid-season.

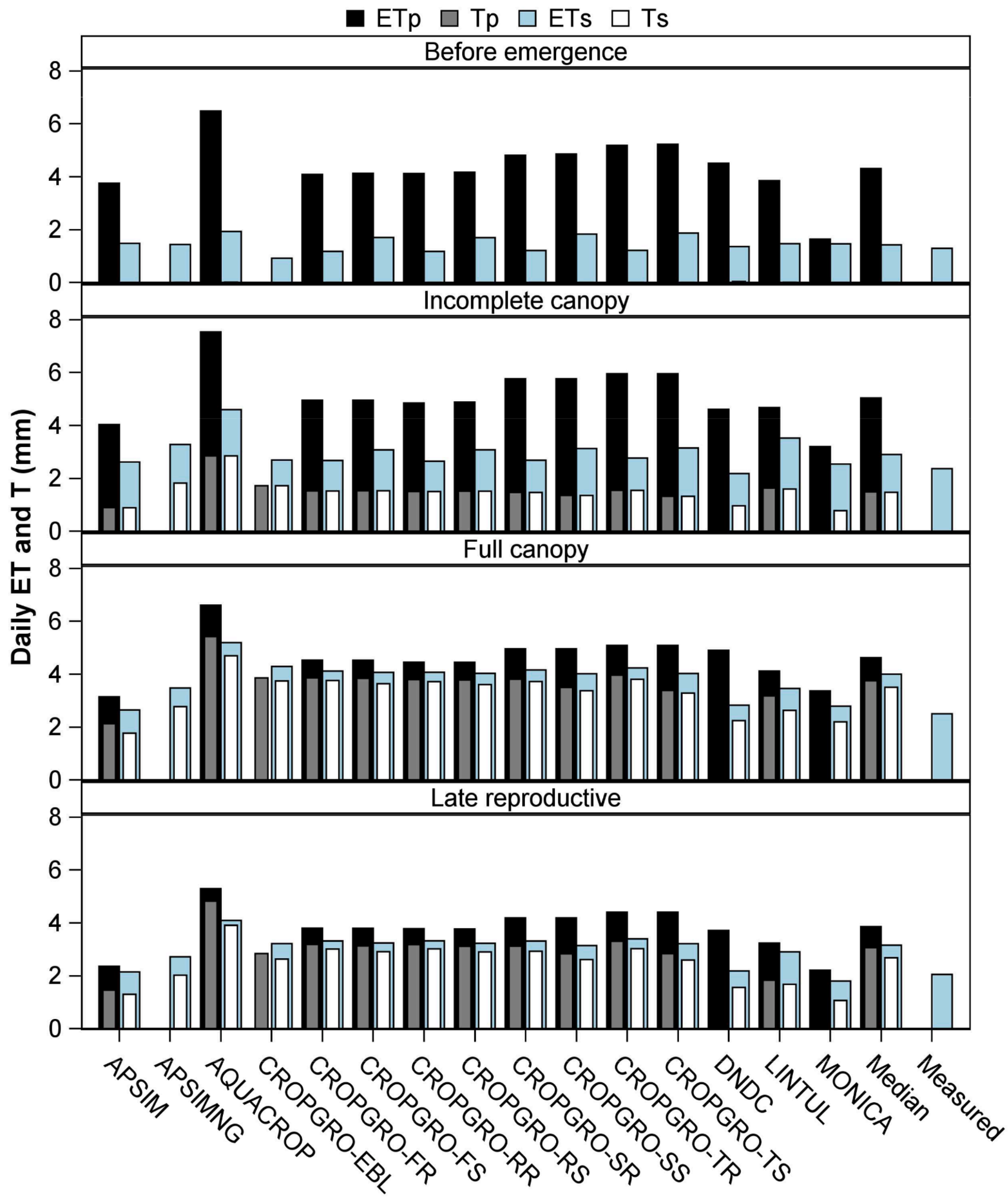

Fig. 7. Mean daily potential ET (ETp), daily evapotranspiration (ETs), daily potential transpiration (Tp), and daily transpiration (Ts) simulated by 15 flavors of soybean models (based on six crop models with two flavors of APSIM and nine flavors of CROPGRO) for periods: (i) before emergence (20 days before emergence to emergence), (ii) incomplete canopy (emergence to day of LAI reaching value of 2.5), (iii) full canopy (day of LAI beyond 2.5 until 30 days later), and (iv) late reproductive (end of full canopy phase to R8).

Modelers were asked to report outputs from all upstream variables used in the computation of daily ETp when possible, based on the current model output files and model structure. Most models reported values for daily ETo. An exception was the CROPGRO group of models that used either first-principles of energy balance (CROPGRO-EBL), estimated ETp directly based on Priestley-Taylor (CROPGRO-RR, CROPGRO-RS), or computed an internal ETo based on the Sau et al. (2004) modification of FAO-56 (CROPGRO-FR, CROPGRO-FS). Across models computing ETo for a short (grass) reference crop, the average daily ETo during the period of full-canopy ranged from 3.8 to 6.0 mm d^{-1} depending on the model (data not shown). For models such as CROPGRO-EBL, daily ETp and upstream variables were not computed given the energy balance approach utilized to simulate ET in the model. However, for other models, some of these variables were involved in the computation of daily ET but were not a regular output for these models.

Overall, there was a large variability in the type and magnitude of upstream variables involved in the computation of daily ETp that is consistent with the findings by Kimball et al. (2023). Despite this variability, the magnitudes of the resulting ETp values were relatively similar across all the models (Fig. 7), except higher ETp for AQUACROP and lower values for MONICA. The results reveal the need to standardize definitions and outputs of variables used in the computation of daily ETp across models, which can facilitate model comparison and model development efforts to improve the simulation of daily ET.

3.6. *Performance of crop models to simulate crop LAI, aboveground biomass, and seed yield*

The accurate simulation of LAI is a key component for the simulation of ET and soil water content, given that most models rely on LAI to compute Ep and Tp from the atmospheric demand or ETp. For instance, CROPGRO computes the amount of solar radiation (MJ m^{-2}) that is absorbed by the foliage and the amount of solar radiation that reaches the soil as a function of the LAI, thus impacting Ep and Tp (Sau et al., 1999; Boote et al., 2008) (Table S1).

AQUACROP uses the percentage of canopy ground cover instead of LAI to partition ETp (Steduto et al., 2009; Khoshravesh et al., 2013). Most other models utilized different approaches to calculate Tp and Ep from ETp, but using a similar conceptual mechanism that relies on LAI (see Supplementary Material, Table S2). It is important to note that measured data on LAI was already provided to modelers in step 1. However, the nRMSE for the simulation of LAI slightly decreased after calibration step 2, compared to step 1 in most models (Fig. 8). The best performance for LAI simulation was obtained by the ensemble median with the lowest nRMSE (28.2%) and absolute bias (2.5) during step 2.

Of interest, the DNDC model was the best-performing model for the simulation of daily ET (Fig. 3), but it was not among the top five models in either of the two steps for the simulation of time-series LAI (Fig. 8a, b). The relatively high nRSME for simulation of LAI in the DNDC model was associated with an underestimation of LAI on average (nBias = −8.6% after step 1, and nBias = −22.4% after step 2) (Fig. 8b). Unlike some of the more physiological-based models, DNDC simulates crop transpiration based on a crop coefficient and a user-defined crop water requirement, while LAI is a product of the daily crop growth but not directly used to compute the partitioning of ETp into Tp and Ep (Table S1). Thus, the overall underestimation of LAI and changes in the simulation of LAI after steps 1 and 2 in the DNDC model had little impact on ETs, which was close to the measured ET for this model. Similar to the DNDC model, the MONICA model also underestimated LAI after step 1 (nBias = −12.9%), and the underestimation intensified after step 2 (nBias= -27.6%). However, MONICA improved the simulation of daily ET in step 2 due to adjustments in Kc across crop developmental stages and in the crop-specific maximum rooting depth (Table S3). The Kc and maximum root depth in the MONICA model are defined as crop coefficients that are calibrated at the cultivar level (Battisti et al., 2017).

The nRMSE for the simulation of the final AGB and final seed yield did not show consistent improvements between calibration steps 1 and 2 for most of the models (Figs. 8d and 9e). However, the limited effect of calibration step 2 on the nRMSE for simulating AGB and seed yield was not a surprise, given that the measured data was already provided to modelers in step 1. After step 2 of calibration, the nRMSE ranged from 8.5% to 42.3 % for AGB and from 1.2% to 28.5% for seed yield. The best performance for the simulation of AGB was obtained by DNDC (nRMSE = 7.1%) after step 1 and CROPGRO-FR (nRMSE = 8.2%) after step 2. For the simulation of seed yield, the best performance was obtained by AQUACROP (nRMSE = 3.3%) after step 1, and DNDC (nRMSE = 3.3%) after step 2.

We analyzed the relationship between the nRMSE for the simulation of daily ET, and the nRMSE for simulation of LAI and AGB (Fig. S6). Unexpectedly, we did not find that the improved simulation of LAI was associated with an improved simulation of daily ET, with the exception of the LINTUL model (r^2=0.75, $p < 0.01$, Fig. S6). Similarly, we did not find an association between the nRMSE for simulation of ABG and daily et (Fig. S6).

Notably, DNDC, despite underestimating LAI and AGB, performed well in simulating daily ET due to its reliance on user-defined crop coefficients rather than physiological parameters, allowing it to decouple biomass accumulation from ETp partitioning. In case of the MONICA model, it improved daily ET simulation after step 2 although it underestimated LAI and AGB, due to adjustments in crop-specific parameters, such as Kc and rooting depth. These results demonstrate that while LAI and AGB are critical inputs for ET partitioning in most models, their influence on daily ET simulation varies depending on the model structure and calibration strategy and can be relatively small.

3.7. Performance of crop models to simulate daily soil water content

Simulated SWC data were evaluated against SWC at the experimental site, measured at different soil depths ranging from 0.05 to 1.10 m (Table 1). To evaluate the models, simulated and measured SWC were

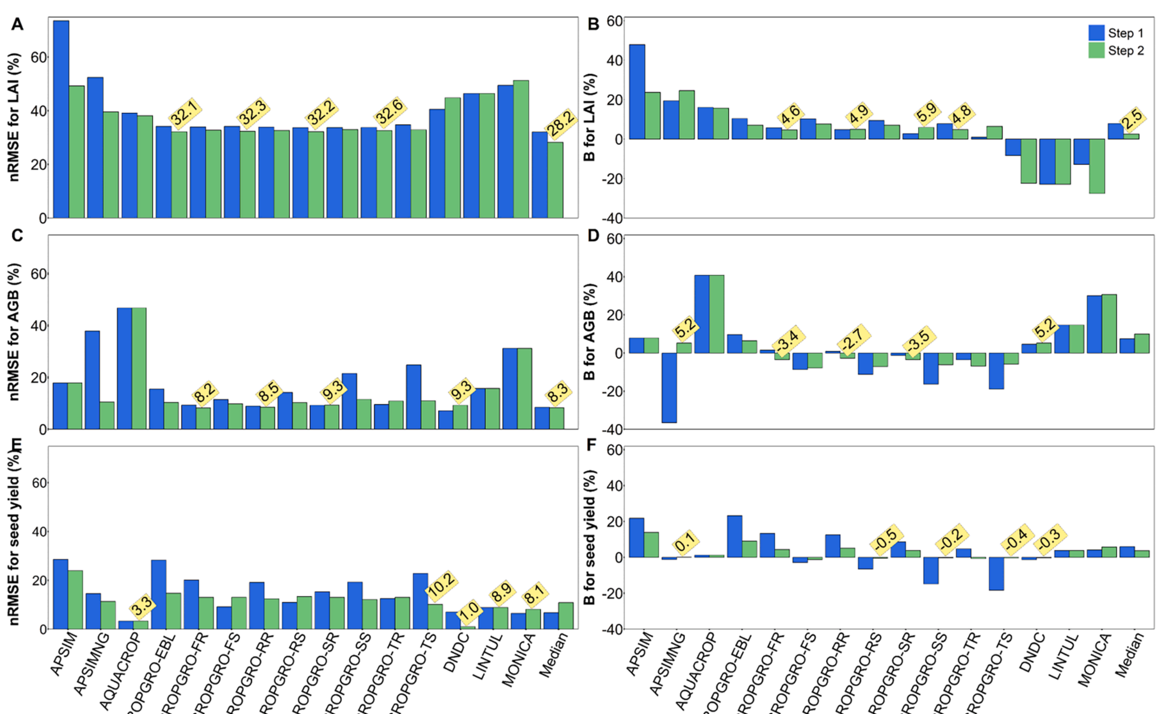

Fig. 8. Normalized root mean square error (nRMSE) for leaf area index [LAI (A)], total aboveground biomass [AGB (C)], and seed yield (E). Normalized bias (nBias) between measured and simulated values for LAI (B), AGB (D), and seed yield (F) over seasons (1997, 1999, 2008, 2016 early, 2016 late, and 2019). Ensemble median values were based on the models APSIM, APSIMNG, AQUACROP, CROPGRO-RR, CROPGRO-EBL, LINTUL, DNDC, and MONICA. The top 5 performing models, after step 2, for nRMSE and B are shown with yellow flags.

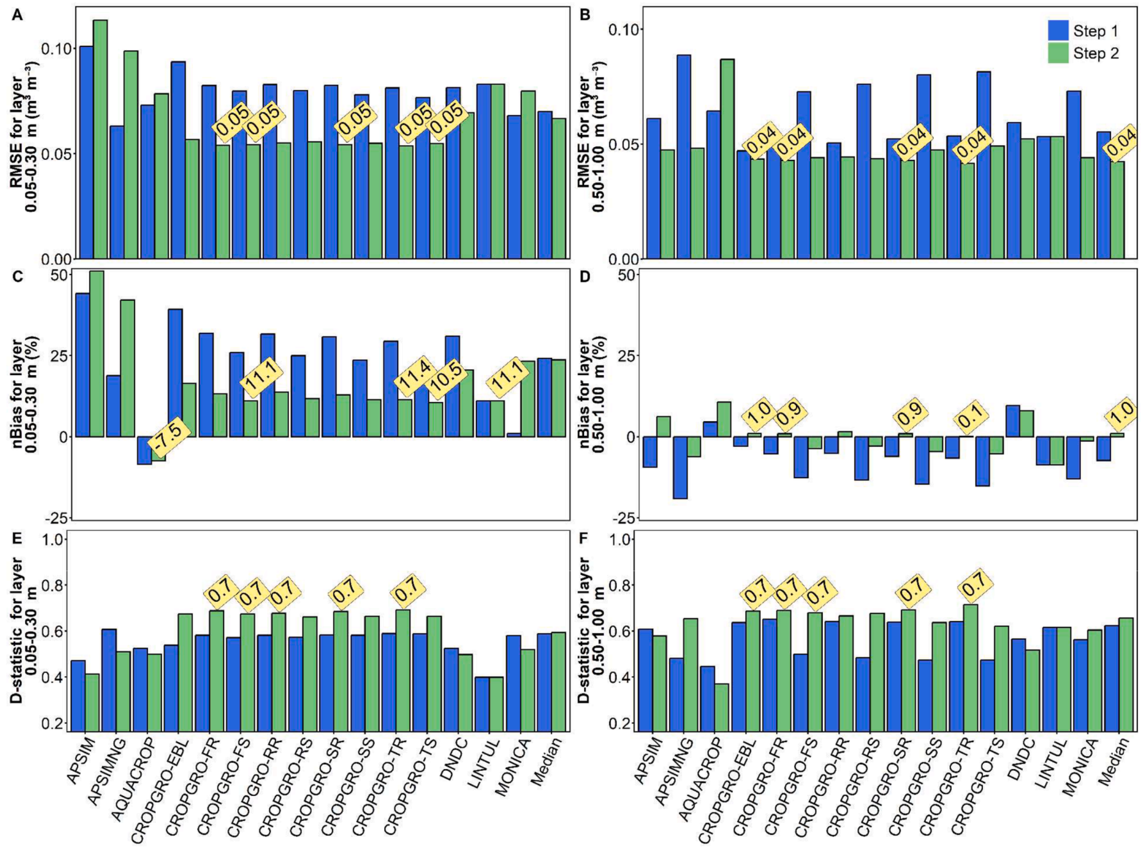

Fig. 9. Root mean squared error (RMSE), D-statistic, and normalized bias (nBias) for time-series volumetric soil water content by model and soil depth. Soil depths were grouped from 0.05 to 0.30 m soil depth (A, C, E) and 0.50 to 1.00 m (B, D, F). Ensemble median values are based on the models APSIM, APSIMNG, AQUACROP, CROPGRO-RR, CROPGRO-EBL, LINTUL, DNDC, and MONICA. The top five performing models, after step 2, for RMSE, B, and D-statistics are shown with yellow flags.

grouped into two soil depth categories (Fig. 9). After calibration step 1, the APSIMNG showed the lowest RMSE for simulation of SWC in the top (RMSE = 0.06 $m^3\ m^{-3}$). After calibration step 2, the CROPGRO group of models showed the lowest RMSE for simulation of SWC in the top (RMSE=0.05 $m^3\ m^{-3}$) and bottom soil layers (RMSE= 0.04 $m^3\ m^{-3}$), with identical performance among flavors CROPGRO-FR, FS, SR, TR, and TS.

Model calibration with measured SWC data improved the simulation of SWC for most models, as indicated by the decrease in RMSE and increase in D-statistic (Fig. 9). Similar to the model performance assessed with the RMSE, the CROPGRO group of models had the best performance for simulation of SWC compared to other models when assessed using the D-Statistic (average of 0.70 across all soil layers and CROPGRO ET approaches). The soil water balance in CROPGRO uses a simple tipping bucket approach for the simulation of volumetric soil water content (Ritchie, 1998). However, the CROPGRO model had been previously evaluated and showed good performance for the simulation of SWC under a wide range of climates, including a tropical environment in Brazil (Figueiredo Moura da Silva et al., 2021), a humid oceanic climate in North-west Spain (Sau et al., 1999), and for cotton grown in a cold desert climate (Li et al., 2019). Different from previously published works, our findings demonstrated that the different methods of soil water evaporation in CROPGRO resulted in similar results in SWC estimates.

It is interesting to note that the models that showed superior performance simulating close to the measured ET data (DNDC, MONICA, and APSIM) were not at the top of the ranking for the simulation of SWC. Overall, most models tended to overestimate SWC in the surface soil layers (0.05 – 0.30 m) and underestimate SWC in relatively deeper soil layers (0.50 – 1.10 m). The DNDC model improved the simulation of both ET and SWC after step 2 relative to calibration with crop growth data only in step 1. The overestimation of SWC in the DNDC model persisted after step 2, with an average bias of 19.3 and 10.9% after steps 1 and 2, respectively. However, it is important to note that runoff and drainage data were not available to examine the entire water budget in this study. In the case of the APSIM and MONICA models, simulated SWC increased across all soil layers after step 2, likely due to the decrease in simulated ET and crop water uptake to better match measured ET data. This caused the simulation of SWC to worsen across all soil layers for APSIM (average bias of 18.7% and 29.6% in steps 1 and 2, respectively), and for the topsoil layers in MONICA (bias of 1.0% and 23.3% for step 1 and 2, respectively) (Fig. 9c,d).

The regression analysis (Figure S6) revealed a positive Pearson correlation between the nRMSE of daily ET and the nRMSE of SWC for the Ensemble (r^2=0.64–0.72, $p < 0.01$ for Step 1 and 2), CROPGRO-RR (r^2=0.73, $p < 0.01$ for Step 1), and LINTUL models (r^2=0.79, $p < 0.01$ for Step 1 and 2). In addition, there was a significant correlation between the nRMSE of daily ET and the nRMSE of SWC when analyzing this relationship with data across all models (r^2=0.38, $p < 0001$ in Step 1). This indicates that better model performance for simulating daily ET was generally associated with more accurate simulations of SWC in the top layers, likely reflecting the role of soil water evaporation and its strong influence on surface ET dynamics. In contrast, the relationship between the nRMSE of daily ET and SWC in deeper soil layers was negative for several models (CROPGRO-EBL, CROPGRO-RR, DNDC, Median), in particular during calibration Step 1 (r^2=−0.87 to −0.97, $p <$ 0.01). The relationship between the nRMSE for daily ET and SWC in bottom layers was also negative when analyzing data across all models (r^2=−0.29, $p < 0.01$). This negative correlation suggests that models prioritizing surface soil water processes to improve ET simulations may have compromised the accuracy of water redistribution to deeper layers.

3.8. General considerations of crop model performance in simulating ET

The main objective of this inter-comparison among soybean models was to quantify uncertainty for the simulation of ET in soybean process-based models after calibration with different levels of measured data. In addition, this study provided an opportunity to evaluate best methods and better parametrize models for the simulation of soybean ET in a humid continental climate and to identify opportunities for further model improvement. Structural and methodological differences among models appear to explain some of the variability observed among models using similar ET methods.

The DNDC model performed better for the simulation of daily ET compared to other models and the ensemble median when only crop growth data were used for calibration. Incorporating observed ET and SWC data during the step 2 in calibration further improved DNDC's simulations of ET and SWC. Although. DNDC had been previously evaluated for simulating ET in this climate (Dutta et al., 2016; Guest et al., 2018; Sansoulet et al., 2014), it did not rank highly for simulating LAI and SWC, which are typically closely associated with transpiration and available water for crop use in models. This lower ranking can be attributed to DNDC's reliance on user-defined crop water use coefficients rather than LAI to estimate transpiration, setting it apart from physiology-based models like CROPGRO, APSIM, and MONICA. The CROPGRO flavors in particular focused on accurately simulating LAI, AGB, and SWC, relying on prior ET parameterizations without recalibrating ET parameters.

When evaluating models for the simulation of SWC, most models tended to overestimate SWC in the surface soil layers (0–0.25 m) and underestimate SWC in deep soil layers (0.5 – 1.0 m). The general overestimation of SWC by the DNDC model may have resulted from an over-calibration of crop water use to prioritize a better fit of model ET estimates to eddy covariance measurements; however, SWC did improve in step 2 as compared to step 1 indicating that other parts of the water budget (*i.e.*, drainage) may also be a factor. Overall, the CROPGRO group of models had the best performance for the simulation of SWC compared to other models, using either the Priestley–Taylor or the FAO-56 Penman–Monteith method for ETo combined with either the Ritchie-Two-Stage or the Suleiman–Ritchie soil water evaporation methods, or using an energy balance approach to compute ET. All the CROPGRO flavors in the study relied on the same approach to simulate soil water balance, using the tipping bucket method (Ritchie, 1998).

We expected that improved model performance for simulation of SWC in some models may contribute to their improved performance for simulation of daily ET. Our results revealed a positive correlation between top-layer SWC and ET across models after Step 1 (r^2=0.38, $p < 0.001$; Fig.S6), which highlights the importance of accurate surface soil water dynamics for simulating actual ET in this humid climate.. Unexpectedly, there was a negative correlation between the nRMSE for daily ET and the nRMSE for SWC in bottom-layers (r^2=0.29, $p < 0.01$, relationship across models). This negative indicates an overcompensation in simulated water uptake or drainage parameters in order to better match measured ET, but that resulted in an underprediction of SWC in deep layers (Fig. 9F).

The overestimation of SWC in the topsoil layer was also previously documented for maize simulations using the STICS model in the same location and soil (CFIA field, Table S1), especially when dry spells occurred, leading to the formation of shrinkage cracks (Saadi et al., 2022). A new method to improve simulation of soil water transport through shrinkage cracks during dry years, named *Improving Shrinkage Cracks Impact* (ISCI), was successful in improving simulation of SWC in the study by Saadi et al. (2022). Approaches that account for soil water transport through shrinkage cracks such as the ISCI method might be an avenue to improve simulation of SWC in other crop models based on the tipping bucket approach.

The CROPGRO flavors provided a unique framework to assess differences in ET methods without the confounding effects of variations in model structure, as these models share identical representations of crop growth and water processes. Comparisons among Priestley-Taylor, FAO-56, and energy balance methods highlighted distinct strengths and limitations of these approaches (Fig.S6). This indicates that the ET approaches within these models played a more critical role in determining ET accuracy than the precision of growth simulations, such as LAI partitioning, particularly during the initial calibration step. The Lintul model was an exception, where better simulation of LAI positively contributed to ET performance (Fig. S4), suggesting differences in model sensitivity to growth variables.

Supplementary material (Table S3) provides details on the parameters modified during calibration (LAI, AGB, and ET), offering insights into the most critical variables for improving ET simulations. It is important to note that this study did not aim to conduct a sensitivity or uncertainty analysis to ET-related parameters as part of the calibration process. This is because most models do not require recalibration of ET parameters, and instead rely on parametrization based on published methods such as FAO-56 and Priestley-Taylor, except in specific cases where adjustments were warranted. Only MONICA, DNDC, APSIM, and APSIMNG modified parameters directly related to ET calculation (Table S3). To improve simulation of daily ET during Step 2, the crop coefficient (Kc) was reduced in MONICA; a soil water evaporation factor was decreased in DNDC; and a factor related to the calculation of saturated vapor pressure (Fv) was decreased in the APSIM models (Table S3).

An unexpected major finding from this study was that most models simulated daily ET values that were consistently higher compared to measured ET through eddy covariance during the crop growing season. The bias between simulated and measured ET became more evident as the crop growing season progressed, with increasing fractions of the ET attributed to crop transpiration. Similar findings were obtained by Crepeau et al. (2021) for simulation of daily ETs in soybean with the STICS model at the same location, but not for maize (Saadi et al., 2022) when using a resistance approach to compute ETs based on Shuttleworth and Wallace (1985). A hypothesis that may partially explain the higher simulated ET compared to measured ET could be due to models in this study not accounting properly for dew formation or incoming water through the night in this humid continental climate that may partially offset total daily ET losses. Dew formation can be a significant contribution to the water budget when the canopy temperature falls to or below the dew temperature of the air in humid climates (Jacobs et al., 2006; Groh et al., 2019), but occurs less often in arid and semiarid regions (Malek et al., 1999). Dew formation can occur when evening radiative cooling causes air water vapor to condense on a surface (Jacobs and Nieveen, 1995), or when water evaporation from the soil is condensed on the crop canopy (Monteith, 1957; Garratt, 1992). Crop models in this study computed a daily evaporative demand based on daily meteorological data that likely do not account properly for these night-time variations in energy fluxes and "negative" ET. In addition, previous studies employing the surface energy balance model from Penman-Monteith in hourly time-steps including night-time found that this approach underestimated dew formation relative to measured data from weight lysimeters (Groh et al., 2018, 2019). Estimates of dew formation and ET during the night with the Penman-Monteith approach improved for a grassland model when the aerodynamic and surface resistance parameters were based on vegetation height observations and the nighttime stomatal resistance was assumed to be zero (Groh et al., 2019).

Similarly, the overestimation of simulated ET could be also due to models not accounting for misty continuous rain that would limit ET during parts of the day, particularly because all the models in our study were run on daily time-steps, except CROPGRO-EBL which runs hourly during the daytime only but with daily weather inputs but not during the night time.

Lastly, the difference between simulated and measured ET data may be partially due to uncertainty in the measured data that may be caused by systematic measurement errors (Moncrieff et al., 1996). The eddy-covariance method can have limitations in estimating gas fluxes during the night under stable atmospheric conditions combined with small gas fluxes and low wind velocity, which make conditions unsuitable for measuring turbulent fluxes (Pattey et al., 2002). For instance, Pattey et al. (2002) highlighted the complexity in quantifying nighttime CO_2 fluxes under the same climate in our study, which had 60% or more of the nights classified as calm during the crop growing season. Fortunately, nighttime ET is expected to be negligible or slightly negative under calm conditions. However, during daytime, eddy flux measurements can be affected occasionally by mesoscale circulation, but the impact of this process on ET measurements cannot be evaluated with one flux tower, as it requires computing ensemble averages to account for larger scale effects. Prior research for an irrigated cotton crop in Bushland, TX indicated that the eddy flux method tended to underestimate ET during the day when compared with mass balance data from weighing lysimeters (Chávez et al. 2009; Alfieri et al., 2011, 2012). Particularly under strongly advective conditions, the eddy flux measurements of latent heat fluxes were lower compared to measurements from large monolithic weighing lysimeters (Alfieri et al., 2012). The results by Alfieri (2012) illustrated the uncertainty in turbulent flux measurements under advective conditions where warm and dry air circulates across irrigated cotton fields. Despite some uncertainty in ET measurements from the eddy covariance method, Foken et al. (2012) in the "Eddy Covariance a practical guide to measurement and data analysis", do not recommend correcting ET data for energy budget closure when turbulent-organized structures are included.

Given the different factors that may explain the general over-estimation of ET by most of these models, it is not possible to conclude with certainty whether crop models had an error simulating ET during the crop growing season, or if there was a possible bias in measurements from eddy covariance in this climate. Thus, the results indicating that most models over-estimate ET must be interpreted with caution since it can be possible that models are not as far off as presented from the results. The models would benefit from reviewing methods for the calculation of daily ETp and further evaluating daily ETs against additional data from warm-summer humid continental climates. We have documented a consistent bias across most of the models for the simulation of soybean ET in this climate. Most published ET methodologies, particularly the FAO-56, were developed in and for arid environments, which differ significantly from the conditions in this study. Our site, characterized by cooler temperatures, frequent rainfall, and lower atmospheric water demand, provides a stark contrast to the environments where these methodologies have been widely evaluated. For instance, AQUACROP, which relies on the well-established FAO-56 methodology, performed poorly in this study, despite its widespread use and success in arid and semi-arid climates (Toumi et al., 2016; Sandhu and Irmak, 2019). This highlights the limitations of such approaches when applied to humid continental climates, emphasizing the need for models to account for unique environmental conditions, such as the cooler and wetter characteristics of this region. In addition, ET methods in these crop models should consider hourly time steps and ET contributions from dew formation at night, which is lacking for daily day-time weather-driven approaches. Nonetheless, all models in this study showed great potential to be applied for simulating carbon and water dynamics, with accurate simulations of the temporal patterns in crop growth, ET, and soil water content.

4. Conclusions

A total of 15 flavors of soybean models (based on six crop models with two flavors of APSIM and nine flavors of DSSAT-CSM-CROPGRO), were compared for their ability to simulate daily and cumulative ET, soybean growth, and soil water content over five growing seasons in Ottawa, Canada. Overall, model-simulated daily ET was consistently higher compared to measured ET values during the crop growing season, particularly during days of relatively high ET. DNDC, a Canadian-developed model with prior experience simulating ET on the site, showed the best performance for simulating daily ET. The MONICA model performed best for the simulation of cumulative ET, which was associated with a direct calibration of Kc coefficients to estimate ET across crop stages. CROPGRO showed the best performance for simulating soil water content. The ensemble median was useful in reducing uncertainty in simulating crop growth, ET, and soil water content, but did not rank first. This study provided an opportunity to better parametrize models for the simulation of ET for a soybean crop and revealed a bias between measured ET from eddy covariance flux towers and that simulated by process-based crop models. This result suggests the need for further investigation in measurement and modeling of ET. In particular, we recommend reviewing approaches in models for computing ETp and Tp and considering ET methods that account for dew formation. Regarding measured ET data from flux systems, we recommend further evaluating whether possible underestimations of ET may occur during the crop growing season, especially under stable conditions and relatively high ET demand.

CRediT authorship contribution statement

Evandro H. Figueiredo Moura da Silva: Writing – original draft, Visualization, Investigation, Formal analysis. **Kritika Kothari:** Writing – review & editing, Visualization, Formal analysis, Data curation. **Elizabeth Pattey:** Writing – review & editing, Methodology, Investigation, Data curation. **Rafael Battisti:** Writing – review & editing, Methodology, Investigation, Conceptualization. **Kenneth J. Boote:** Writing – review & editing, Methodology, Funding acquisition, Conceptualization. **Sotirios V. Archontoulis:** Writing – review & editing, Investigation. **Santiago Vianna Cuadra:** Writing – review & editing, Investigation. **Babacar Faye:** Writing – review & editing, Investigation. **Brian Grant:** Writing – review & editing, Investigation. **Gerrit Hoogenboom:** Writing – review & editing, Investigation. **Qi Jing:** Writing – review & editing, Investigation. **Fábio R. Marin:** Writing – review & editing, Investigation. **Claas Nendel:** Writing – review & editing, Investigation. **Budong Qian:** Writing – review & editing, Investigation. **Ward Smith:** Writing – review & editing, Investigation. **Amit Kumar Srivastava:** Writing – review & editing, Investigation. **Kelly R. Thorp:** Writing – review & editing, Investigation. **Nilson A. Vieira Junior:** Writing – review & editing, Investigation. **Montserrat Salmerón:** Writing – review & editing, Project administration, Methodology, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Montserrat Salmeron reports financial support was provided by National Institute of Food and Agriculture. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors are grateful to the National Institute of Food and Agriculture, United States Department of Agriculture (USDA-NIFA Grant # 2019-67019-29470) for funding this study.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.agrformet.2025.110463.

Data availability

Data will be made available on request.

References

Alfieri, J.G., Kustas, W.P., Prueger, J.H., Hipps, L.E., Chávez, J.L., French, A.N., Evett, S. R., 2011. Intercomparison of nine micrometeorological stations during the BEAREX08 field campaign. J. Atmos. Ocean. Technol. 28 (11), 1390–1406.

Alfieri, J.G., Kustas, W.P., Prueger, J.H., Chavez, J.L., Evett, S.R., Neale, C.M., Anderson, M.C., Hipps, L.E., Copeland, K.S., Howell, T.A., French, A.N., Dulaney, W., McKee, L., 2012. A comparison of the eddy covariance and lysimetry-based measurements of the surface energy fluxes during BEAREX08. In: Remote Sensing and Hydrology Symposium, pp. 215–218.

Allen, R.G., Fisher, D.K., 1991. Direct load cell-based weighing lysimeter system. In: Grouting in Geotechnical Engineering. ASCE, pp. 114–124.

Allen, R.G., Pereira, L.S., Raes, D., Smith, M., 1998. Crop Evapotranspiration-Guidelines for Computing Crop Water requirements-FAO Irrigation and Drainage Paper 56, 300. FAO, Rome, p. D05109.

Asseng, S., Ewert, F., Rosenzweig, C., Jones, J.W., Hatfield, J.L., Ruane, A.C., Boote, K.J., Thorburn, P.J., Rötter, R.P., Cammarano, D., Brisson, N., Basso, B., Martre, P., Aggarwal, P.K., Angulo, C., Bertuzzi, P., Biernath, C., Challinor, A.J., Doltra, J., Gayler, S., Goldberg, R., Grant, R., Heng, L., Hooker, J., Hunt, L.A., Ingwersen, J., Izaurralde, R.C., Kersebaum, K.C., Müller, C., Naresh Kumar, S., Nendel, C., O'Leary, G., Olesen, J.E., Osborne, T.M., Palosuo, T., Priesack, E., Ripoche, D., Semenov, M.A., Shcherbak, I., Steduto, P., Stöckle, C., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Travasso, M., Waha, K., Wallach, D., White, J.W., Williams, J.R., Wolf, J., 2013. Uncertainty in simulating wheat yields under climate change. Nat. Clim. Chang. 3 (9), 827–832.

Bassu, S., Brisson, N., Durand, J.L., Boote, K., Lizaso, J., Jones, J.W., Rosenzweig, C., Ruane, A., Adam, M., Baron, C., Basso, B., Biernath, C., Boogaard, H., Conijn, S., Corbeels, M., Deryng, D., Sanctis, G.D., Gayler, S., Grassini, P., Hatfield, J., Hoek, S., Izaurralde, C., Jongschaap, R., Kemanian, A.R., Kersebaum, K.C., Kim, S., Naresh S K., Makowski, D., Muller, C., Nendel, C., Priesack, E., Pravia, M.V., Sau, F., Shcherbak, I., Tao, F., Teixeira, E., Timlin, D., Waha, K., 2014. How do various maize crop models vary in their responses to climate change factors? Glob. Chang. Biol. 20 (7), 2301–2320.

Battisti, R., Sentelhas, P.C., Boote, K.J., 2017. Inter-comparison of performance of soybean crop simulation models and their ensemble in southern Brazil. Field Crops Res. 200, 28–37.

Battisti, R., Sentelhas, P.C., Boote, K.J., 2018. Sensitivity and requirement of improvements of four soybean crop simulation models for climate change studies in Southern Brazil. Int. J. Biometeorol. 62, 823–832.

Boote, K.J., Jones, J.W., Hoogenboom, G., Pickering, N.B., 1998. The CROPGRO model for grain legumes. Understanding Options for Agricultural Production. Springer, Dordrecht, pp. 99–128.

Boote, K.J., Sau, F., Hoogenboom, G., Jones, J.W., 2008. Experience with water balance, evapotranspiration, and simulations of water stress effects in the CROPGRO model. In: Ahuja, L.R., Reddy, V.R., Saseendran, S.A., Yu, Qiang (Eds.), Response of Crops to Limited Water: Understanding and Modeling Water Stress Effects On Plant Growth Processes, V.1. Advances in Agricultural Systems Modeling - ASA-CSSA-SSSA, Madison, WI, pp. 59–103.

Bowen, I.S., 1926. The ratio of heat losses by conduction and by evaporation from any water surface. Phys. Rev. 27 (6), 779.

Cammarano, D., Rötter, R.P., Asseng, S., Ewert, F., Wallach, W., Martre, P., Hatfield, J.L., Jones, J.W., Rosenzweig, C., Ruane, A.C., Boote, K.J., Thorburn, P.J., Kersebaum, K. C., Aggarwal, P.K., Angulo, C., Basso, B., Bertuzzi, P., Biernath, C., Brisson, N., Challinor, A.J., Doltra, J., Gayler, S., Goldberg, R., Heng, L., Hooker, J.E., Hunt, L.A., Ingwersen, J., Izaurraldez, R.C., Müller, C., Kumar, S.N., Nendel, C., O'Leary, G., Olesen, J.E., Osborne, T.M., Priesack, E., Ripoche, D., Steduto, P., Stöckle, C.O., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Travasso, M., Waha, K., White, J.W., Wolf, J., 2016. Uncertainty of wheat water use: simulated patterns and sensitivity to temperature and CO2. Field Crops Res. 198, 80–92.

Chartres, C.J., Noble, A., 2015. Sustainable intensification: overcoming land and water constraints on food production. Food Secur. 7 (2), 235–245.

Chatterjee, A., Anapalli, S.S., 2023. Comparison of cropping system models for simulation of soybean evapotranspiration with eddy covariance measurements in a humid subtropical environment. Water 15 (17), 3078.

Chávez, J.L., Howell, T.A., Copeland, K.S., 2009. Evaluating eddy covariance cotton ET measurements in an advective environment with large weighing lysimeters. Irrig. Sci. 28 (1), 35–50.

Cohen, I., Huang, Y., Chen, J., Benesty, J., Benesty, J., Chen, J., Cohen, I., 2009. Pearson correlation coefficient. In: Noise Reduction in Speech Processing, pp. 1–4.

Crépeau, M., Jégo, G., Morissette, R., Pattey, E., Morrison, M.J., 2021. Predictions of soybean harvest index evolution and evapotranspiration using STICS crop model. Agron. J. 113 (4), 3281–3298.

Cuadra, S.V., Kimball, B.A., Boote, K.J., Suyker, A.E., Pickering, N., 2021. Energy balance in the DSSAT-CSM-CROPGRO model. Agric. For. Meteorol. 297, 108241.

DeJonge, K.C., Thorp, K.R., 2017. Implementing standardized reference evapotranspiration and dual crop coefficient approach in the DSSAT cropping system model. Trans. ASABe 60 (6), 1965–1981.

Dias, H.B., Cuadra, S.V., Boote, K.J., Lamparelli, R.A.C., Figueiredo, G.K.D.A., Suyker, A. E., Magalhães, P.S.G., Hoogenboom, G., 2023. Coupling the CSM-CROPGRO-Soybean crop model with the ECOSMOS Ecosystem Model–An evaluation with data from an AmeriFlux site. Agric. For. Meteorol. 342, 109697.

Din, M.S.U., Mubeen, M., Hussain, S., Ahmad, A., Hussain, N., Ali, M.A., Sabagh, A.E., Elsabagh, M., Shah, G.M., Qaisrani, S.A., Tahir, M., Javeed, H.M.R., Anwar-ul-Haq, M., Ali, M., Nasim, W., 2022. World nations priorities on climate change and food security. Building Climate Resilience in Agriculture. Springer, Cham, pp. 365–384.

Dutta, B., Grant, B.B., Congreves, K.A., Smith, W.N., Wagner-Riddle, C., VanderZaag, A. C., Tenuta, M., Desjardins, R.L., 2018. Characterizing effects of management practices,snow cover, and soil texture on soil temperature: model development in DNDC. Biosyst. Eng. 168, 54–72.

Dutta, B., Smith, W.N., Grant, B.B., Pattey, E., Desjardins, R.L., Li, C., 2016. Model development in DNDC for the prediction of evapotranspiration and water use in temperate field cropping systems. Environ. Model. Softw. 80, 9–25.

Environment and Climate Change Canada, 2022. Canadian Climate Normals. (accessed 15 November 2022). https://climate.weather.gc.ca/.

Figueiredo Moura da Silva, E.H., Boote, K.J., Hoogenboom, G., Gonçalves, A.O., Junior, A.S.A., Marin, F.R., 2021. Performance of the CSM-CROPGRO-soybean in simulating soybean growth and development and the soil water balance for a tropical environment. Agric. Water. Manag. 252, 106929.

Figueiredo Moura da Silva, E.H., Hoogenboom, G., Boote, K.J., Gonçalves, A.O., Marin, F. R., 2022. Predicting soybean evapotranspiration and crop water productivity for a tropical environment using the CSM-CROPGRO-Soybean model. Agric. For. Meteorol. 323, 109075.

Figueiredo Moura da Silva, E.F.M., La Menza, N.C., Munareto, G.G., Zanon, A.J., Carvalho, K.S., Marin, F.R., 2023. Soybean seed protein concentration is limited by nitrogen supply in tropical and subtropical environments in Brazil. J. Agric. Sci. 161 (2), 279–290.

Fleisher, D.H., Condori, B., Quiroz, R., Alva, A., Asseng, S., Barreda, C., Bindi, M., Boote, K.J., Ferrise, R., Franke, A.C., Panamanna, M., Govindakrishnan, D.H., Hoogenboom, G., Kumar, S.N., Merante, P., Nendel, C., Olesen, J.E., Parker, P.S., Raes, D., Raymundo, R., Ruane, A.C., Stockle, C., Supit, I., Vanuytrecht, E., Wolf, J., Woli, P., 2017. A potato model intercomparison across varying climates and productivity levels. Glob. Chang. Biol. 23 (3), 1258–1281.

Chap.4: 85-131 Foken, T., Leuning, R., Oncley, S.R., Murder, M., Aubinet, M., 2012. Corrections and data quality control. In: Aubinet, M., Vesala, T, Papale, D. (Eds.), Eddy Covariance - A Practical Guide to Measurement and Data Analysis. Series, 1st ed. Springer Atmospheric Sciences, p. 270. 2012XISBN 978-94-007-2350-4.

Garratt, J.R., 1992. Extreme maximum land surface temperatures. J. Appl. Meteorol. Climatol. 31 (9), 1096–1105.

Gauch, H.G., Hwang, J.G., Fick, G.W., 2003. Model evaluation by comparison of model-based predictions and measured values. Agron. J. 95 (6), 1442–1446.

Gijsman, A.J., Jagtap, S.S., Jones, J.W., 2002. Wading through a swamp of complete confusion: how to choose a method for estimating soil water retention parameters for crop models. Eur. J. Agron. 18 (1–2), 77–106.

Groh, J., Slawitsch, V., Herndl, M., Graf, A., Vereecken, H., Pütz, T., 2018. Determining dew and hoar frost formation for a low mountain range and alpine grassland site by weighable lysimeter. J. Hydrol. 563, 372–381.

Groh, J., Pütz, T., Gerke, H.H., Vanderborght, J., Vereecken, H., 2019. Quantification and prediction of nighttime evapotranspiration for two distinct grassland ecosystems. Water Resour. Res. 55 (4), 2961–2975.

Guarin, J.R., Asseng, S., 2022. Improving wheat production and breeding strategies using crop models. Wheat Improvement. Springer, Cham, pp. 573–591.

Guest, G., Smith, W., Grant, B., McConkey, B., Chipanshi, A., Reid, K., Kroebel, R., Martel, M., Desjardins, R., VanderZaag, A., Pattey, E., Glenn, A., Wilson, H., Balde, H., Wagner-Riddle, C., Drury, C.F., Fuller, K., Hayashi, M., Reynolds, D., 2018. Comparing the performance of the DNDC, Holos, and VSMB models for predicting the water partitioning of various crops and sites across Canada. Can. J. Soil Sci. 98 (2), 212–231.

Hoogenboom, G., 2000. Contribution of agrometeorology to the simulation of crop production and its applications. Agric. For. Meteorol. 103 (1–2), 137–157. https://doi.org/10.1016/S0168-1923(00)00108-8.

Hoogenboom, G., Porter, C.H., Boote, K.J., Shelia, V., Wilkens, P.W., Singh, U., White, J. W., Asseng, S., Lizaso, J.I., Moreno, P., Pavan, W., Ogoshi, R., Hunt, A., Tsuji, G.Y., Jones, J.W., 2019. The DSSAT crop modeling ecosystem. Advances in Crop Modelling For a Sustainable Agriculture. Burleigh Dodds Science Publishing, pp. 173–216.

Hoogeveen, J., Faurès, J.M., Van de Giessen, N., 2009. Increased biofuel production in the coming decade: to what extent will it affect global freshwater resources? Irrig. Drain. 58 (S1), S148–S160.

Holzworth, D., Huth, N.I., Fainges, J., Brown, H., Zurcher, E., Cichota, R., Verrall, S., Herrmann, N.I., Zheng, B., Snow, V., 2018. APSIM Next Generation: overcoming challenges in modernising a farming systems model. Environ. Model. Softw. 103, 43–51.

Hsiao, T.C., Heng, L., Steduto, P., Rojas-Lara, B., Raes, D., Fereres, E., 2009. AquaCrop—The FAO crop model to simulate yield response to water: III. Parameterization and testing for maize. Agron. J. 101 (3), 448–459.

Jacobs, A.F., Heusinkveld, B.G., Wichink Kruit, R.J., Berkowicz, S.M., 2006. Contribution of dew to the water budget of a grassland area in the Netherlands. Water Resour. Res. 42 (3).

Jacobs, A.F., Nieveen, J.P., 1995. Formation of dew and the drying process within crop canopies. Meteorol. Appl. 2 (3), 249–256.

Jones, J.W., Hoogenboom, G., Porter, C.H., Boote, K.J., Batchelor, W.D., Hunt, L.A., Wilkens, P.W., Singh, U., Gijsman, A.J., Ritchie, J.T., 2003. The DSSAT cropping system model. Eur. J. Agron. 18 (3), 235–265.

Keating, B.A., Carberry, P.S., Hammer, G.L., Probert, M.E., Robertson, M.J., Holzworth, D., Huth, N.I., Hargreaves, J.N.G., Meinke, H., Hochman, Z., McLean, G., Verburg, K., Snow, V., Dimes, J.P., Silburn, M., Wang, E., Brown, S., Bristow, K.L., Asseng, S., Chapman, S., McCown, R.L., Freebairn, D.M., Smith, C.J., 2003. An overview of APSIM, a model designed for farming systems simulation. Eur. J. Agron. 18 (3–4), 267–288.
Keating, B.A., Thorburn, P.J., 2018. Modelling crops and cropping systems—Evolving purpose, practice and prospects. Eur. J. Agron. 100, 163–176.
Khoshravesh, M., Mostafazadeh-Fard, B., Heidarpour, M., Kiani, A.R., 2013. AquaCrop model simulation under different irrigation water and nitrogen strategies. Water Sci. Technol. 67 (1), 232–238.
Kimball, B.A., Boote, K.J., Hatfield, J.L., Ahuja, L.R., Stockle, C., Archontoulis, S., Baron, C., Basso, B., Bertuzzi, P., Constantin, J., Deryng, D., Dumont, B., Durand, J. L., Ewert, F., Gaiser, T., Gayler, S., Hoffmann, M.P., Jing, Q., Soo-Hyung, K., Lizaso, J., Moulin, S., Nandel, C., Parker, P., Palosuo, T., Priesack, E., Zhiming, Q., Srivastava, A., Stella, T., Tao, F., Thorp, K.R., Timlin, D., Twine, T.E., Webber, H., Willaume, M., Williams, K., 2019. Simulation of maize evapotranspiration: an inter-comparison among 29 maize models. Agric. For. Meteorol. 271, 264–284.
Kimball, B.A., Thorp, K.R., Boote, K.J., Stockle, C., Suyker, A.E., Evett, S.R., Brauer, D.K., Coyle, G.G., Copeland, K.S., Marek, G.W., Colaizzi, P.D., Acutis, M., Alimagham, S., Archontoulis, S., Babacar, F., Barcza, Z., Basso, B., Bertuzzi, P., Constantin, J., Miglionari, M.A., Dumont, B., Durand, J.L., Fodor, N., Gaiser, T., Garofalo, P., Gayler, S., Giglio, L., Grant, R., Guan, K., Hoogenboom, G., Jiang, Q., Kim, S.H., Kisekka, I., Lizaso, J., Masia, S., Meng, H., Mereu, V., Mukhtar, A., Perego, A., Peng, B., Priesack, E., Qi, Z., Vakhtang, S., Snyder, R., Soltani, A., Spano, D., Srivastava, A., Thomson, A., Timlin, D., Trabucco, A., Webber, H., Weber, T., Willaume, M., Williams, K., Lann, M.V.D., Ventrella, S., Viswanathan, M., Xu, X., Zhou, W., 2023. Simulation of evapotranspiration and yield of maize: an Inter-comparison among 41 maize models. Agric. For. Meteorol. 333, 109396.
Koeppen, W., 1948. Koeppen Climatologia: Con Un Estudio de Los Climas De La Tierra. Fondo de Cultura Econômica, Mexico City, Mexico, p. 478.
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., Silva, F.A.M., 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 Jr, N.A., Slone, S., Salmerón, M., 2022. Are soybean models ready for climate change food impact assessments? Eur. J. Agron. 135, 126482.
Kothari, K., Battisti, R., Boote, K.J., Archontoulis, S.V., Confalone, A., Constantin, J., Cuadra, S., Debaeke, P., Faye, B., Grant, B., Hoogenboom, G., Jing, Q., van der Lann, M., Silva, F.A.M., Marin, F.R., Nehbandani, A., Nendel, C., Purcell, L.C., Qian, B., Ruane, A.C., Salmerón, M., 2024. Evaluating differences among crop models in simulating soybean in-season growth. Field Crops Res. 309, 109306.
Li, C., Frolking, S., Frolking, T.A., 1992. A model of nitrous oxide evolution from soil driven by rainfall events: 1. Model structure and sensitivity. J. Geophys. Res. Atmos. 97 (D9), 9759–9776.
Li, T., Hasegawa, T., Yin, X., Zhu, Y., Boote, K., Adam, M., Bregaglio, S., Buis, S., Confalonieri, R., Fumoto, T., Gaydon, D., Marcaida III, M., Nakagawa, H., Oriol, P., Ruane, A.C., Ruget, F., Singh, B., Singh, U., Tang, L., Tao, L., Wilkens, P., Yoshida, H., Zhang, Z., Bouman, B., 2015. Uncertainties in predicting rice yield by current crop models under a wide range of climatic conditions. Glob. Chang. Biol. 21 (3), 1328–1341.
Li, M., Du, Y., Zhang, F., Bai, Y., Fan, J., Zhang, J., Chen, S., 2019. Simulation of cotton growth and soil water content under film-mulched drip irrigation using modified CSM-CROPGRO-cotton model. Agric. Water. Manag. 218, 124–138.
Loague, K., Green, R.E., 1991. Statistical and graphical methods for evaluating solute transport models: overview and application. J. Contam. Hydrol. 7 (1–2), 51–73.
Malek, E., McCurdy, G., Giles, B., 1999. Dew contribution to the annual water balances in semi-arid desert valleys. J. Arid Environ. 42 (2), 71–80.
Marin, F.R., Thorburn, P.J., Nassif, D.S., Costa, L.G., 2015. Sugarcane model intercomparison: structural differences and uncertainties under current and potential future climates. Environ. Model. Softw. 72, 372–386.
Mizutani, K., Yamanoi, K., Ikeda, T., Watanabe, T., 1997. Applicability of the eddy correlation method to measure sensible heat transfer to forest under rainfall conditions. Agric. For. Meteorol. 86 (3–4), 193–203.
Moncrieff, J.B., Malhi, Y., Leuning, R., 1996. The propagation of errors in long-term measurements of land-atmosphere fluxes of carbon and water. Glob. Chang. Biol. 2 (3), 231–240.
Monteith, J.L., 1957. Dew. Q. J. R. Meteorol. Soc. 83 (357), 322–341.
Ndehedehe, C., 2022. Impacts of Water Resources Development on Hydrology. Satellite Remote Sensing of Terrestrial Hydrology. Springer, Cham, pp. 389–437.
Nendel, C., Berg, M., Kersebaum, K.C., Mirschel, W., Specka, X., Wegehenkel, M., Wenkel, K.O., Wieland, R., 2011. The MONICA model: testing predictability for crop growth, soil moisture and nitrogen dynamics. Ecol. Modell. 222 (9), 1614–1625. https://doi.org/10.1016/j.ecolmodel.2011.02.018.
Pattey, E., Edwards, G., Strachan, I.B., Desjardins, R.L., Kaharabata, S., Wagner Riddle, C., 2006. Towards standards for measuring greenhouse gas flux from agricultural fields using instrumented towers. Can. J. Soil Sci. 86, 373–400.
Pattey, E., Strachan, I.B., Boisvert, J.B., Desjardins, R.L., McLaughlin, N.B., 2001. Detecting effects of nitrogen rate and weather on corn growth using micrometeorological and hyperspectral reflectance measurements. Agric. For. Meteorol. 108, 85–99.
Pattey, E., Strachan, I.B., Desjardins, R.L., Massheder, J., 2002. Measuring nighttime CO2 flux over terrestrial ecosystems using eddy covariance and nocturnal boundary layer methods. Agric. For. Meteorol. 113 (1–4), 145–158.
Penman, H.L., 1948. Natural evaporation from open water, bare soil and grass. Proc. R. Soc. Lond. A Math. Phys. Sci. 193 (1032), 120–145.
Priestley, C.H.B., Taylor, R.J., 1972. On the assessment of surface heat flux and evaporation using large-scale parameters. Mon. Weather. Rev. 100 (2), 81–92.
Purcell, L.C., King, C.A., 1996. Drought and nitrogen source effects on nitrogen nutrition, seed growth, and yield in soybean. J. Plant Nutr. 19 (6), 969–993.
Puy, A., Lo Piano, S., Saltelli, A., 2020. Current models underestimate future irrigated areas. Geophys. Res. Lett. 47 (8), e2020GL087360.
Puy, A., Borgonovo, E., Lo Piano, S., Levin, S.A., Saltelli, A., 2021. Irrigated areas drive irrigation water withdrawals. Nat. Commun. 12 (1), 1–12.
Raes, D., Steduto, P., Hsiao, T.C., Fereres, E., 2009. AquaCrop—The FAO crop model to simulate yield response to water: II. Main algorithms and software description. Agron. J. 101 (3), 438–447.
Ritchie, J.T., 1972. Model for predicting evaporation from a row crop with incomplete cover. Water Resour. Res. 8 (5), 1204–1213.
Ritchie, J.T., 1998. Soil water balance and plant water stress. Understanding Options for Agricultural Production. Springer, Dordrecht, pp. 41–54.
Ritchie, J.T., Porter, C.H., Judge, J., Jones, J.W., Suleiman, A.A., 2009. Extension of an existing model for soil water evaporation and redistribution under high water content conditions. Soil Sci. Soc. Am. J. 73 (3), 792–801.
Robertson, M.J., Carberry, P.S., 1998. Simulating growth and development of soybean in APSIM.
Rosenzweig, C., Jones, J.W., Hatfield, J.L., Ruane, A.C., Boote, K.J., Thorburn, P., Antle, J.M., Nelson, G.C., Porter, C., Janssen, S., Asseng, S., Basso, B., Ewert, F., Wallach, D., Baigorria, G., Winter, J.M., 2013. The agricultural model intercomparison and improvement project (AgMIP): protocols and pilot studies. Agric. For. Meteorol. 170, 166–182.
Rosenzweig, C., Solecki, W., 2014. Hurricane Sandy and adaptation pathways in New York: lessons from a first-responder city. Glob. Environ. Chang. 28, 395–408.
Rosenzweig, C., Mutter, C.Z., Ruane, A.C., Contreras, E.M., Boote, K.J., Valdivia, R.O., Houtkamp, J., Homann-Kee Tui, S., Ahmad, A., Subash, N., Vellingiri, G., Nedumaran, S., 2021. AgMIP regional integrated assessments: high-level findings, methods, tools, and studies (2012–2017). In: Handbook of Climate Change and Agroecosystems, 5. World Scientific Publishing, pp. 123–142.
Saadi, S., Pattey, E., Champagne, E., Jégo, G., 2022. Prediction of rainfed corn evapotranspiration and soil moisture using the STICS crop model in eastern Canada. Field Crop Rese. 284, 108664.
Sandhu, R., Irmak, S., 2019. Performance of AquaCrop model in simulating maize growth, yield, and evapotranspiration under rainfed, limited and full irrigation. Agric. Water. Manage 223, 105687.
Sansoulet, J., Pattey, E., Kröbel, R., Grant, B., Smith, W., Jégo, G., Desjardins, R.L., Tremblay, G., 2014. Comparing the performance of the STICS, DNDC, and DayCent models for predicting N uptake and biomass of spring wheat in Eastern Canada. Field Crops Res. 156, 135–150.
Sau, F., Boote, K.J., Ruíz-Nogueira, B., 1999. Evaluation and improvement of CROPGRO-soybean model for a cool environment in Galicia, northwest Spain. Field Crops Res. 61 (3), 273–291.
Sau, F., Boote, K.J., Bostick, W.M., Jones, J.W., Mínguez, M.I., 2004. Testing and improving evapotranspiration and soil water balance of the DSSAT crop models. Agron J 96 (5).
Saxton, K.E., Rawls, W., Romberger, J.S., Papendick, R.I., 1986. Estimating generalized soil-water characteristics from texture. Soil science society of America Journal 50 (4), 1031–1036.
Shuttleworth, W.J., Wallace, J.S., 1985. Evaporation from sparse crops-an energy combination theory. Q. J. R. Meteorol. Soc. 111 (469), 839–855.
Siebert, S., Döll, P., Hoogeveen, J., Faures, J.M., Frenken, K., Feick, S., 2005. Development and validation of the global map of irrigation areas. Hydrol. Earth. Syst. Sci. 9 (5), 535–547.
Smith, W., Grant, B., Qi, Z., He, W., VanderZaag, A., Drury, C.F., Helmers, M., 2020. Development of the DNDC model to improve soil hydrology and incorporate mechanistic tile drainage: a comparative analysis with RZWQM2. Environ. Model. Softw. 123, 104577.
Snow, V.O., Huth, N.I., 2004. The APSIM–Micromet module. The APSIM–Micromet Module, 12848, 21.
Soycanada, 2022. At a Glance: canadian Soybean Industry. Retrieved from https://soycanada.ca/statistics/at-a-glance/.
Steduto, P., Hsiao, T.C., Raes, D., Fereres, E., 2009. AquaCrop—The FAO crop model to simulate yield response to water: I. Concepts and underlying principles. Agron. J. 101 (3), 426–437.
Suleiman, A.A., Ritchie, J.T., 2003. Modeling soil water redistribution during second-stage evaporation. Soil Sci. Soc. Am. J. 67 (2), 377–386.
Suleiman, A.A., Ritchie, J.T., 2004. Modifications to the DSSAT vertical drainage model for more accurate soil water dynamics estimation. Soil Sci. 169 (11), 745–757.
Thorp, K.R., Marek, G.W., DeJonge, K.C., Evett, S.R., 2020. Comparison of evapotranspiration methods in the DSSAT Cropping System Model: II. Alborithm performance. Comput. Electron. Agric. 177, 105679.
Toumi, J., Er-Raki, S., Ezzahar, J., Khabba, S., Jarlan, L., Chehbouni, A., 2016. Performance assessment of AquaCrop model for estimating evapotranspiration, soil water content and grain yield of winter wheat in Tensift Al Haouz (Morocco): application to irrigation management. Agric. Water. Manag. 163, 219–235.
USDA US Department of Agriculture, National Agricultural Statistics Service, 2022. Census of Agriculture: united States Summary and State Data. Volume 1, Geographic Area Series, Part 51. AC–17-A-51.
Wang, E., Martre, P., Ewert, F., Zhao, Z., Maiorano, A., Rötter, R.P., Kimball, B.A., Ottman, M.J., Wall, G.W., White, J.W., Reynolds, M.P., Alderman, P.D., Aggarwal, P. K., Anothai, J., Basso, B., Biernath Cammarano, D., Challinor, A.J., De Sanctis, G., Doltra, J., Fereres, E., Garcia-Vila, M., Gayler, S., Hoogenboom, G., Hunt, L.A., Izaurralde, R.C., Jabloun, M., Jones, C.D., Kersebaum, K.C., Koehler, A.-K.,

Müller, C., Liu, L., Kumar, S.N., Nendel, C., O'Leary, G., Olesen, J.E., Palosuo, T., Priesack, E., Rezaei, E.E., Ripoche, D., Ruane, A.C., Semenov, M.A., Shcherbak, I., Stöckle, C., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Thorburn, P., Waha, K., Wallach, D., Wang, Z., Wolf, J., Zhu, Y., Asseng, S., 2017. The uncertainty of seed yield projectionsis reduced by improved temperature response functions. Nat. Plants 3 (1702), 1–11.

Wang, C., Linderholm, H.W., Song, Y., Wang, F., Liu, Y., Tian, J., Xu, J., Song, Y., Ren, G., 2020. Impacts of drought on maize and soybean production in northeast China during the past five decades. Int. J. Environ. Res. Public Health 17 (7), 2459.

Wang, B., Jägermeyr, J., O'Leary, G.J., Wallach, D., Ruane, A.C., Feng, P., Linchao, L., De Liu, L., Waters, C., Qiang, Y., Asseng, S., Rosenzweig, C., 2024. Pathways to identify and reduce uncertainties in agricultural climate impact assessments. Nat. Food 5 (7), 550–556.

Westhoek, H., Rood, T., Van den Berg, M., Janse, J., Nijdam, D., Reudink, M., Stehfest, E., Lesschen, J.P., Oenema, O., Woltjer, G.B., 2011. The Protein puzzle: the Consumption and Production of meat, Dairy and Fish in the European Union (No. 500166001). Netherlands Environmental Assessment Agency.

White, J.W., Hoogenboom, G., Kimball, B.A., Wall, G.W., 2011. Methodologies for simulating impacts of climate change on crop production. Field Crops Res. 124 (3), 357–368.

Willmott, C.J., Ackleson, S.G., Davis, R.E., Feddema, J.J., Klink, K.M., Legates, D.R., O'Donnell, J., Rowe, C.M., 1985. Statistics for the evaluation and comparison of models. J. Geophys. Res. Oceans 90 (C5), 8995–9005.

Wolf, J., 2012. LINTUL5: simple Generic Model for Simulation of Crop Growth under Potential, Water Limited and Nitrogen. Phosphorus and Potassium Limited Conditions.

Wu, G., Fanzo, J., Miller, D.D., Pingali, P., Post, M., Steiner, J.L., Thalacker-Mercer, A.E., 2014. Production and supply of high-quality food protein for human consumption: sustainability, challenges, and innovations. Ann. N. Y. Acad. Sci. 1321 (1), 1–19.

Wu, B., Tian, F., Zhang, M., Piao, S., Zeng, H., Zhu, W., Liu, Elnashar, A., Lu, Y., 2022. Quantifying global agricultural water appropriation with data derived from earth observations. J. Clean. Prod. 358, 131891.

Yasin, M., Ahmad, A., Khaliq, T., Habib-ur-Rahman, M., Niaz, S., Gaiser, T., Ghafoor, I., Hassan, H.S.U., Qasim, M., Hoogenboom, G., 2022. Climate change impact uncertainty assessment and adaptations for sustainable maize production using multi-crop and climate models. Environ. Sci. Pollut. Res. 29 (13), 18967–18988.

Zhang, Y., Niu, H., 2016. The development of the DNDC plant growth sub-model and the application of DNDC in agriculture: a review. Agric. Ecosyst. Environ. 230, 271–282.