Agricultural and Forest Meteorology 333 (2023) 109396

Contents lists available at ScienceDirect

Agricultural and Forest Meteorology

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

Simulation of evapotranspiration and yield of maize: An inter-comparison among 41 maize models

Bruce A. Kimball[a,*], Kelly R. Thorp[a], Kenneth J. Boote[b], Claudio Stockle[c], Andrew E. Suyker[d], Steven R. Evett[e], David K. Brauer[e], Gwen G. Coyle[e], Karen S. Copeland[e], Gary W. Marek[e], Paul D. Colaizzi[e], Marco Acutis[f], Seyyedmajid Alimagham[g], Sotirios Archontoulis[h,1], Faye Babacar[i], Zoltán Barcza[j,k], Bruno Basso[l], Patrick Bertuzzi[m], Julie Constantin[n], Massimiliano De Antoni Migliorati[o], Benjamin Dumont[p], Jean-Louis Durand[q], Nándor Fodor[k,r], Thomas Gaiser[s], Pasquale Garofalo[t], Sebastian Gayler[u], Luisa Giglio[t], Robert Grant[v], Kaiyu Guan[w], Gerrit Hoogenboom[b], Qianjing Jiang[x], Soo-Hyung Kim[y], Isaya Kisekka[z], Jon Lizaso[A], Sara Masia[B], Huimin Meng[C], Valentina Mereu[D], Ahmed Mukhtar[E], Alessia Perego[f], Bin Peng[w], Eckart Priesack[F], Zhiming Qi[x], Vakhtang Shelia[b], Richard Snyder[G], Afshin Soltani[g], Donatella Spano[D], Amit Srivastava[s], Aimee Thomson[H], Dennis Timlin[I], Antonio Trabucco[D], Heidi Webber[J], Tobias Weber[u], Magali Willaume[n], Karina Williams[K,L], Michael van der Laan[H], Domenico Ventrella[t], Michelle Viswanathan[u], Xu Xu[C], Wang Zhou[w]

[a] *U.S. Arid-Land Agricultural Research Center, USDA-ARS, Maricopa, AZ 85138*
[b] *University of Florida, Agricultural and Biological Engineering, Frazier Rogers Hall, Gainesville, Florida 32611-0570, USA*
[c] *Biological Systems Engineering, Washington State University, 1935 E. Grimes Way, PO Box 646120, Washington State University, Pullman WA 99164-6120*
[d] *School of Natural Resources, University of Nebraska-Lincoln, Lincoln, Nebraska, USA*
[e] *Conservation and Production Research Laboratory, USDA-ARS, Bushland, Texas, USA*
[f] *Department of Agricultural and Environmental Sciences, University of Milan, via Celoria 2 - 20133, Milan, Italy*
[g] *Agronomy Group, Gorgan Univ. of Agric. Sci. & Natur. Resour., Gorgan 49138-15739 Iran*
[h] *Iowa State University, Department of Agronomy, Ames, Iowa 50010*
[i] *Institut de recherche pour le développement (IRD) ESPACE-DEV, F-34093 Montpellier Cedex, France*
[j] *ELTE Eötvös Loránd University, Department of Meteorology, H-1192 Budapest, Hungary*
[k] *Czech University of Life Sciences Prague, Faculty of Forestry and Wood Sciences, 165 21 Prague, Czech Republic*
[l] *Michigan State University, Dept. Geological Sciences and W.K. Kellogg Biological Station, 288 Farm Ln, 307 Natural Science Bldg., East Lansing, MI, 48824*
[m] *US1116 AgroClim, INRAE centre de recherche Provence-Alpes-Côte d'Azur, 228, route de l'Aérodrome, CS 40 509, Domaine Saint Paul, Site Agroparc, 84914 Avignon Cedex 9, France*
[n] *AGIR, Université de Toulouse, INRAE, INPT, INP- EI PURPAN, 24 Chemin de Borde Rouge - Auzeville CS 52627, CastanetTolosan, France*
[o] *Queensland Department of Environment & Science, Queensland, Australia*
[p] *ULiege-GxABT, University of Liege - Gembloux Agro-Bio Tech, TERRA Teaching and research centre, Plant Science Axis / Crop Science Lab., B-5030 Gembloux, Belgium*
[q] *Unité de Recherches Pluridisciplinaire Prairies et Plantes Fourragères, INRAE, 86 600 Lusignan, France*
[r] *Agricultural Institute, Centre for Agricultural Research, H-2462 Martonvásár, Brunszvik u. 2., Hungary*
[s] *Institute of Crop Science and Resource Conservation, University of Bonn, Katzenburgweg 5, D-53115 Bonn, Germany*
[t] *Agricultural Research and Economics, Agriculture and Environment Research Center, CREA-AA, Via Celso Ulpiani 5, 70125 BARI BA, Italy*
[u] *Universität Hohenheim, Institute of Soil Science and Land Evaluation, Biogeophysics, Emil-Wolff-Str. 27, D-70593 Stuttgart, Germany*
[v] *Department of Renewable Resources, University of Alberta, Edmonton, Alberta, Canada T6G 2E3*
[w] *College of Agricultural, Consumer and Environmental Sciences (ACES), University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA*

Abbreviations: ASCE, American Society of Civil Engineers; DAP, days after planting; E, soil water evaporation; Ep, potential soil water evaporation; Es, simulated soil water evaporation; ET, evapotranspiration; ETo, "short" reference evapotranspiration based on 12-cm-tall grass; ETp, potential evapotranspiration; ETr, "tall" reference evapotranspiration based on 50-cm-tall alfalfa; ETs, simulated evapotranspiration; LAI, leaf area index; nRMSE, normalized root mean square error; MESA, mid-elevation sprinkler application; P-T, Priestley-Taylor; SDI, subsurface drip irrigation; T, transpiration; Tp, potential canopy transpiration; Ts, simulated canopy transpiration.

[1] 1-520-316-6369

* Corresponding author at: U.S. Arid-Land Agricultural Research Center, USDA, Agricultural Research Service, 21881 North Cardon Lane, Maricopa, Arizona 85018, USA.

E-mail address: bruce.kimball@usda.gov (B.A. Kimball).

https://doi.org/10.1016/j.agrformet.2023.109396
Received 26 August 2022; Received in revised form 8 February 2023; Accepted 24 February 2023
Available online 10 March 2023

[x] Department of Bioresource Engineering, Macdonald Campus, McGill University, 1-024 Macdonald-Steward Hall, Sanite-Anne-de-Bellevue, QC, Canada H9X 3V9
[y] 25School of Environmental and Forest Sciences, University of Washington, Seattle, WA 98195
[z] Agricultural Water Management and Irrigation Engineering; University of California Davis; Departments of Land, Air, and Water Resources and of Biological and Agricultural Engineering; One Shields Avenue; PES 1110; Davis, CA 95616-5270, USA
[A] Technical University of Madrid (UPM), Dept. Producción Agraria-CEIGRAM, Ciudad Universitaria, 28040 Madrid, Spain
[B] Land and Water Management Department, IHE Delft Institute for Water Education, Delft, The Netherlands
[C] Center of Agricultural Water Research, China Agricultural University, Beijing, China
[D] Fondazione CMCC - Euro-Mediterranean Centre on Climate Change, Impacts on Agriculture, Forests and Ecosystem Services division (IAFES), Sassari, Italy
[E] Department of Agronomy, PMAS Arid Agriculture University, Rawalpindi, Pakistan and Swedish University of Agricultural Sciences, Umea Sweden
[F] Helmholtz Center Munich, Institute of Biochemical Plant Pathology, Ingolstaedter Landstr. 1, 85764 Neuherberg, Germany
[G] University of California Davis
[H] University of Pretoria, Pretoria, South Africa
[I] Crop Systems and Global Change Research Unit, USDA-ARS, Beltsville, MD
[J] Leibniz Centre for Agricultural Landscape Research (ZALF), Mucheberg 15374, Germany
[K] Hadley Centre, FitzRoy, Road Exeter Devon EX1 3PB, United Kingdom
[L] Global Systems Institute, University of Exeter, North Park Road, Exeter, EX4 4QE, UK

ARTICLE INFO

Keywords:
Maize
Simulation
Evapotranspiration
Water use
Crop models
Yield

ABSTRACT

Accurate simulation of crop water use (evapotranspiration, ET) can help crop growth models to assess the likely effects of climate change on future crop productivity, as well as being an aid for irrigation scheduling for today's growers. To determine how well maize (*Zea mays* L.) growth models can simulate ET, an initial inter-comparison study was conducted in 2019 under the umbrella of AgMIP (Agricultural Model Inter-Comparison and Improvement Project). Herein, we present results of a second inter-comparison study of 41 maize models that was conducted using more comprehensive datasets from two additional sites - Mead, Nebraska, USA and Bushland, Texas, USA. There were 20 treatment-years with varying irrigation levels over multiple seasons at both sites. ET was measured using eddy covariance at Mead and using large weighing lysimeters at Bushland. A wide range in ET rates was simulated among the models, yet several generally were able to simulate ET rates adequately. The ensemble median values were generally close to the observations, but a few of the models sometimes performed better than the median. Many of the models that did well at simulating ET for the Mead site did poorly for drier, windy days at the Bushland site, suggesting they need to improve how they handle humidity and wind. Additional variability came from the approaches used to simulate soil water evaporation. Fortunately, several models were identified that did well at simulating soil water evaporation, canopy transpiration, biomass accumulation, and grain yield. These models were older and have been widely used, which suggests that a larger number of users have tested these models over a wider range of conditions leading to their improvement. These revelations of the better approaches are leading to model improvements and more accurate simulations of ET.

1. Introduction

Crop growth models are a useful management aid for today's farmers, as well as being a tool to forecast the likely effects of climate change on future agricultural productivity and irrigation water requirements. For both tasks they need to be accurate. Therefore, in a major effort to improve their accuracy and reliability, modeling groups within the Agricultural Model Inter-comparison and Improvement Project (AgMIP; https://agmip.org/) have been inter-comparing multiple models against each other and against field datasets with varying CO_2, temperature, nitrogen fertilizer, and water supply [wheat (*Triticum aestivum* L.; Asseng et al., 2013, 2015; Cammarano et al., 2016; Liu et al., 2016; Maiorano et al., 2017; Wang et al., 2017), maize (*Zea mays* L.; Bassu et al., 2014; Durand et al., 2018; Kimball et al., 2019), rice (*Oryza sativa* L.; Li et al., 2015; Hasegawa et al., 2017), and potato (*Solanum tuberosum* L.; Fleisher et al., 2017)].

As discussed by Kimball et al. (2019), only a few comparisons among methods or models to simulate ET have been done previously. Sau et al. (2004) evaluated several ET options with the CROPGRO Faba bean (*Vicia faba* L.) model, by careful comparison to soil water balance, and found that the FAO-56 option (Allen et al., 1998) had a root mean square error (RMSE) that was 20% smaller than the Priestley-Taylor option (Priestley and Taylor, 1972) and 48% smaller than the FAO-24 option (Doorenbos and Pruitt, 1985). In an inter-comparison of water use among 16 wheat models at four sites around the world, Cammarano et al. (2016) found the coefficient of variation was only 22.5% among models and sites. In contrast, in an inter-comparison among 23 maize models, Bassu et al. (2014) found a very large range of simulated values of ET among the models, including −10 to +30% variations in the ET response to doubled CO_2 concentration (720 μmol/mol). However, there were no observations of ET or water use in the dataset chosen for that study, so there was no standard for comparison. Therefore, Kimball et al. (2019) conducted their study using eight seasons of data from Ames, Iowa, USA for which eddy covariance measurements of ET were available. Like Bassu et al. (2014), they also found simulated ET values varied by a factor of two among the maize models. Surprisingly, among the models with closest agreement to observations, some were quite simple (e.g., no simulation of biomass) and some were quite complex (e.g., full energy balance), so it was difficult to determine which approaches were generally best and should be adopted by the poorer performing models. Nevertheless, there were several cases in which different ET methods were tested within the same family of crop models, and comparisons among these methods clearly revealed some approaches that were better than others.

However, there were some issues with the Ames dataset (Kimball et al., 2019). For example, in 2012, an infamous year for drought in the Midwest, observed ET and crop yield were higher than in other years. Further analysis led to the strong suspicion that there was a water table present to provide additional water besides the sparse rainfall, yet deep soil water measurements were lacking to confirm the suspicion. Therefore, it was decided to repeat the study of Kimball et al. (2019) with more robust datasets.

Two such datasets were identified, one from the University of Nebraska at Mead, Nebraska, USA (41.165°N, 96.470°W, 362 m), which is close to the 100th meridian typically used to divide the humid East from the arid West, thus placing it within the U.S. "corn belt." There were six seasons of maize from irrigated and rainfed fields (12 treatment-years) with ET determined using eddy covariance. The second was collected by the USDA, Agricultural Research Service, Conservation and Production Research Laboratory (CPRL), Bushland, Texas, USA

(35.183°N, 102.100°W, 1170 m), which is a more arid region where maize is mostly grown with irrigation, and where winds are commonly higher. They measured ET using large weighing lysimeters. They grew maize for two seasons with MESA (mid-elevation sprinkler application) at 100% and 75% replacement of soil water and in near-duplicate SDI (sub-surface drip irritation) fields at 100% (8 treatment-years). A total of 41 models participated in this second round of maize ET simulation inter-comparisons (Tables 1, S1), and again the primary objective was to identify the approaches that were most accurate for simulating ET, i.e., had the lowest RMSE compared to the observations. Besides ET, other objectives were to test the models' abilities to simulate LAI, biomass, grain yield, soil moisture, and soil temperature. By "approaches" we mean the methods used by the models to simulate ET or other processes, i.e., FAO-56 (Allen et al., 1998) versus Priestley-Taylor (1972), etc.

2. Materials and methods

2.1. Observed data

2.1.1. University of Nebraska, Mead, Nebraska, USA

One set of field data came from the University of Nebraska Agricultural Research and Development Center near Mead, Nebraska, USA (http://csp.unl.edu/public/). The soils were deep silty clay loams of Yutan (fine-silty, mixed, superactive, mesic Mollic Hapludalfs), Tomek (fine, smectitic, mesic Pachic Argialbolls), Filbert (fine, smectitic, mesic Vertic Argialbolls), and Filmore (fine, smectitic, mesic Vertic Argialbolls). The eddy covariance technique was used to determine ET of maize and soybean (*Glycine* max) in alternate years, as well as fluxes of sensible heat and CO_2. Additional details can be found in Suyker and Verma (2008, 2009) and Suyker et al. (2004, 2005). Briefly, fluxes of latent heat, sensible heat, and momentum were determined using data from the following sensors at each site: an omnidirectional 3D sonic anemometer (Model R3: Gill Instruments Ltd., Lymington, UK) and an open-path infrared CO_2/H_2O gas analyzing system (Model LI7500: Li-Cor Inc., Lincoln, NE).

The instruments were deployed near the centers of the fields, and the fetch was about 400 m in all directions. The eddy covariance sensors were mounted 5.5 m above the ground. Fluxes were corrected for inadequate sensor frequency, and they were also adjusted for the variation in air density due to the transfer of water vapor and sensible heat. Air temperature and relative humidity (Humitter50Y, Vaisala, Helsinki, FIN), soil heat flux at 0.06 m (Radiation and Energy Balance Systems, Inc., Seattle, WA), and net radiation at 5.5 m (CNR1, Kipp and Zonen Ltd., Delft, NLD) were also measured. Missing data due to sensor malfunction, power outages, unfavorable weather, etc. (approximately 15–20% per year), were estimated using an approach that combined measurement, interpolation, and empirical data synthesis. When hourly values were missing (day or night), the latent heat values were estimated as a function of available energy. Linear regressions between latent heat and available energy were determined (separately for dry and wet conditions) for sliding 3-day intervals, and these estimates were used to fill in missing flux values.

To check closure of the energy balance, the sum of latent and sensible heat fluxes ($\lambda E + H$) measured by eddy covariance were plotted against the sum of R_n (net radiation) + four storage terms, determined by other methods (e.g., Suyker and Verma, 2008). Linear regressions were calculated between the hourly values of $H + \lambda E$ and $R_n + G$ at the study sites (excluding winter months and periods with rain and irrigation). Here $G = G_s$ (soil heat storage) + G_c (canopy heat storage) + G_m (heat stored in the mulch) + G_p (energy used in photosynthesis). The regression slopes averaged 0.89 ± 0.08, implying a fairly good closure of the energy balance.

We used values of daily ET flux, called observed-ET for 2003, 2005, 2007, 2009, 2011, and 2013 from the US-Ne2 (41.165° N, 96.470° W, 362 m; http://ameriflux.lbl.gov/sites/siteinfo/US-Ne2) irrigated maize-soybean rotation field and from the US-Ne3 (41.180° N, 96.440° W, 363 m; http://ameriflux.lbl.gov/sites/siteinfo/US-Ne3) rainfed maize-soybean rotation field. Conservation tillage practices were used, so plant residues were not ploughed into the soil, and the soil surface was generally partially covered with prior soybean crop residue. Both sites are part of the Ameriflux (https://ameriflux.lbl.gov/sites) U.S. surface gas flux observation system, and the two sites are within 1.6 km of each other. The cultivars were Pn33B51, Pn33G66, Pn33H26, Pn33T57, DK_61–72, and DK_62–98 used in 2003, 2005, 2007, 2009, 2011, and 2013, respectively. The irrigated crops were planted on 14 May, 2 May, 1 May, 21 April, 17 May, and 30 May, and the rainfed crops on 13 May, 26 April, 2 May, 22 April, 2 May, and 13 May in 2003, 2005, 2007, 2009, 2011, and 2013, respectively. Destructive measurements of green leaf area index (LAI) and biomass were made approximately bi-monthly during the growing season.

2.1.2. USDA, agricultural research service, conservation and production research laboratory, Bushland, Texas, USA

Maize was grown in 2013 and 2016 at the USDA-ARS Conservation and Production Research Laboratory (https://www.ars.usda.gov/plains-area/bushland-tx/cprl/), Bushland, Texas (35.18° N, 102.10° W, 1170 m above MSL) on a gently sloping (<0.3%) Pullman soil (fine, mixed, superactive, thermic Torrertic Paleustoll). Additional details and data are provided by Evett et al. (2019, 2020, 2022). Four 4.4 ha fields, approximately square in shape and adjacent to each other, each contained a large (3 $m \times 3$ m in surface area, 2.3 m deep) precision weighing lysimeter in the center. The lysimeters contained undisturbed cores of the Pullman soil obtained on site, and they had an accuracy of 0.04 mm water depth equivalent or better (Evett et al., 2012; Marek et al., 1988). The fields and their associated lysimeters were designated NE, SE, NW, and SW according to the inter-cardinal directions. The NE and SE lysimeters and fields were irrigated by subsurface drip irrigation (SDI), and the NW and SW lysimeters and fields were irrigated by mid-elevation sprinkler application (MESA) using a ten-span linear-move system described by Evett et al. (2019). Adaptation of SDI for the NE and SE weighing lysimeters was described by Evett et al. (2018a). A 109-day drought-tolerant variety (Pioneer 1151AM AquaMax, ≤80% Bt) was planted on 16–17 May 2013 under MESA irrigation, on 22–23 May 2013 in the SDI fields and on 10–11 May 2016 in all fields. These are typical dates for maize planting in the region. Crops were managed and fertilized for high grain yield, as detailed by Evett et al. (2019). In each field, destructive subsampling for leaf area index and biomass occurred in replicate plots periodically during the season, and plant height and row width were measured at the same times. Maize harvests were on 15 October 2013 and on 13 and 17 October 2016.

Soil water content was sensed at center depths of 0.10 to 2.30 m in 0.20 m increments in each of eight access tubes in the field around each lysimeter and in two access tubes in each lysimeter (to 1.90 m depth) on a weekly basis, unless prevented by wet field conditions, using a field-calibrated neutron probe and depth-control stand (Evett et al., 2008). Once the crop was established, irrigations were scheduled weekly to replenish the soil water in the top 1.5 m of the profile to field capacity (i.e., replenishing 100% of crop ET), except for one MESA field where irrigations were 75% of full crop ET after crop establishment. As explained by Evett et al. (2019), the MESA 75% deficit irrigation treatment was established to complete a previous longer-term deficit irrigation study. In some cases, two or even three irrigations were required in a week to replenish the water used by the crop. Irrigations by sprinkler and by SDI typically did not occur on the same day. Neutron probe readings were delayed until the soil surface was dry enough to walk on. The soil profile in early 2013 was quite dry, and SDI preplant irrigation and SDI irrigation immediately after planting were required to plant and germinate the crop. This resulted in a full soil profile in the SDI fields by the time neutron probe sensing began, while crop germination with MESA irrigation was accomplished with less frequent irrigations that did not penetrate to the 1.5 m depth. Irrigations in the 100% SDI and MESA fields maintained the soil water depletion to less than the management-allowed depletion level throughout the season. In 2016, the soil profile was much wetter following a wet winter, and no preplant irrigation and less irrigation immediately after planting were needed. Again, irrigations in the 100% SDI and MESA fields kept soil water depletion to less than the management-allowed depletion level.

Evapotranspiration (ET) was determined on 5 min, 15 min, and daily bases using data analyses and quality control procedures described by Marek et al. (2014) and Evett et al. (2019). Fifteen-minute-average weather data were output from the research weather station of the USDA-ARS Soil and Water Management Research Unit at Bushland, Texas located immediately east of the lysimeter fields. The weather station instrumentation and data quality assurance and control procedures were applied as described by Evett et al. (2018b).

2.2. Modeling methodology

2.2.1. Model list

The simulations were conducted by 20 modeling groups from around the world with 41 models completing the inter-comparison (Table 1). Details about each model are presented in supplementary Table S1. However, as can be seen from the names (Tables 1, S1), in some cases there were several "flavors" of different simulation methods tested within the same model family that were chosen by the user at run time. The biggest example is that of the DSSAT family (Hoogenboom et al., 2019a,b; Jones et al., 2003) of the Cropping System Model (CSM) within which both the CSM-CERES-Maize and CSM-IXIM-Maize (hereafter simply called CERES and IXIM) modules were run. Both calculate a value called potential evapotranspiration, ETp, which was done using four methods: (1) FAO-56 (Allen et al., 1998), (2) Priestley-Taylor (1972), (3) the ASCE Standardized Reference Evapotranspiration Equation (R.G. Allen et al., 2005) for 12-cm grass (short crop), and (4) the ASCE Equation for 50-cm alfalfa (tall crop; *Medicago sativa* L.) with FAO-56 dual crop coefficients for maize (Table S1). Within these eight combinations, two E methods for calculating soil water evaporation were tested: "Ritchie" (Ritchie, 1972) and "Suleiman" (Suleiman and Ritchie, 2003, 2004). In addition, within the CERES-FAO-56 and CERES-Priestley-Taylor combinations, E was also computed using Hydrus (Šimůnek et al., 1998, 2008; Shelia et al., 2018), in which soil water moves based on potential gradients. Thus, there were a total of 18 (2 models x 4 ETp methods x 2 soil E methods + 2 Hydrus) DSSAT flavors. Within the DSSAT flavors, model calibrations though Phase 4 were aimed at the best statistics [lowest RMSE, and highest D-statistic (Willmont, 1982)] for growth, grain yield, ET, and soil water variables, averaged over four ET options (two ET by two E methods) in order to minimize bias. The ASCE and Hydrus ET options were not included in this process because the methods were not part of the DSSAT V4.7 release, so they were at a slight disadvantage because they were not independently calibrated. Nevertheless, the resulting cultivar coefficients were consistently used among all the DSSAT simulations.

In addition, Expert-N had GECROS and SPASS flavors, STICS had KETP and ETP_SW flavors, and MAIZSIM had daily and hourly flavors.

2.2.2. Simulation protocol

The study was conducted in four phases:

1 "Blind phase." The modelers were sent key input data about soils, weather, and management (planting dates, irrigations, fertilizer applications, etc.) information. They also received anthesis and maturity dates, but no other information about plant growth, grain yield, or water use.
2 "Potential or non-stressed growth phase". The modelers were sent time-series leaf area index (LAI) and biomass observations, as well as final grain yields for all the non-water-stress treatments (only irrigated for Mead; only 100% irrigation for Bushland)
3 "Non-stress ET phase". The modelers were sent all ET, soil water, and soil temperature for the non-water-stress treatments (only irrigated for Mead; only 100% irrigation for Bushland)
4 "All phase". In this final phase, the modelers were provided with all LAI, biomass, grain yield, ET, soil moisture, soil temperature etc. data for all treatment-years.

Table 1
List of models and their acronyms. (For details about the evapotranspiration aspects of each, see Supplementary Table S1: List of Models Plus Their Simulation Characteristics and Comparisons of Soil Moisture Simulations).

Acronym	Model Name	Reference
AHC	Agro-Hydrological & chemical & Crop sys. simulator	Xu et al., 2018
AMSW	APSIM-SOILWAT	Keating et al., 2003
AQCP	AquaCrop	Allen et al., 1998
AQY	Aqyield	Constantin et al., 2015
ARMO	ARMOSA	Perego et al., 2013
BIOM	Biome-BGCMuSo 6.0.2	Hidy et al., 2016
CS	CropSyst4	Stöckle et al., 2003
DACT	DayCent-CABBI	Moore et al., 2020
DCAR	DSSAT CSM-CERES-Maize ASCE-Alfalfa Ritchie	DeJonge and Thorp, 2017
DCAS	DSSAT CSM-CERES-Maize ASCE-Alfalfa Suleiman	DeJonge and Thorp, 2017
DCFH	DSSAT CSM-CERES-Maize FAO-56 Hydrus	2018
DCFR	DSSAT CSM-CERES-Maize FAO-56 Ritchie	Sau et al., 2004
DCFS	DSSAT CSM-CERES-Maize FAO-56 Suleiman	Sau et al., 2004
DCGR	DSSAT CSM-CERES-Maize ASCE-Grass Ritchie	DeJonge and Thorp, 2017
DCGS	DSSAT CSM-CERES-Maize ASCE-Grass Suleiman	DeJonge and Thorp, 2017
DCPH	DSSAT CSM-CERES-Maize Priestley-Taylor Hydrus	2018
DCPR	DSSAT CSM-CERES-Maize Priestley-Taylor Ritchie	Sau et al., 2004
DCPS	DSSAT CSM-CERES-Maize Priestley-Taylor Suleiman	Sau et al., 2004
DIAR	DSSAT CSM-IXIM-Maize ASCE-Alfalfa Ritchie	DeJonge and Thorp, 2017
DIAS	DSSAT CSM-IXIM-Maize ASCE-Alfalfa Suleiman	DeJonge and Thorp, 2017
DIFR	DSSAT CSM-IXIM-Maize FAO-56 Ritchie	Sau et al., 2004
DIFS	DSSAT CSM-IXIM-Maize FAO-56 Suleiman	Sau et al., 2004
DIGR	DSSAT CSM-IXIM-Maize ASCE-Grass Ritchie	DeJonge and Thorp, 2017
DIGS	DSSAT CSM-IXIM-Maize ASCE-Grass Suleiman	DeJonge and Thorp, 2017
DIPR	DSSAT CSM-IXIM-Maize Priestley-Taylor Ritchie	Sau et al., 2004
DIPS	DSSAT CSM-IXIM-Maize Priestley-Taylor Suleiman	Sau et al., 2004
ECOS	ecosys	Grant and Flanagan, 2007
JUL	JULES	Best et al., 2011
L5SH	L5-SLIM-H	Wolf, 2012
MZD	MAIZSIM Daily	Yang et al., 2009
MZH	MAIZSIM Hourly	Yang et al., 2009
SLUS	SALUS	Basso and Ritchie, 2015
SLFT	SIMPLACE LINTUL5 FAO56 SLIM3 CanopyT	Wolf, 2012
SMET	SIMETAW#	Mancosu et al., 2016
SSMi	SSM-iCROP	Soltani and Sinclair, 2012
STCK	STICS_KETP	Brisson et al., 2003
STSW	STICS_ETP_ SW	Brisson et al., 2003
SWB	SWB	Annandale et al., 1999
TMOD	Test Model	
XNGM	Expert-N - GECROS	Priesack et al., 2006
XNSM	Expert-N - SPASS	Priesack et al., 2006

The modelers were told to start their simulations on day-of-year 91, so there would be time for equilibration of soil moisture and soil temperature. They were also provided "initial" soil water content profiles, but the number of days before planting at which soil moisture was determined varied widely from season to season.

2.2.3. Methods for evaluating model performance

Correlation coefficients (r), D statistics (Willmott, 1982), root mean squared errors between observed and simulated values (RMSE), normalized root mean squared errors (nRMSE), average differences, as well as mean squared deviations (MSD), standard bias (SB), non-unity of slopes (NU), and lack of correlations (LC) following Gauch et al. (2003), are all presented as Supplementary Statistical Data for Phase 4. Also included are slopes and intercepts of regressions of observed on simulated data, along with corresponding graphs for each model and analyzed parameter.

Herein, we chose to present the nRMSE results calculated using:

$$\mathrm{nRMSE} = \left\{ \left[\mathrm{n}^{-1} \Sigma (\mathrm{P_i} - \mathrm{O_i})^2 \right]^{0.5} \right\} \bar{\mathrm{O}}^{-1}$$

where n = number of observations, P_i and O_i are the simulated and observed ith value pair, and $\bar{O}$ is the observed mean. Normalizing with $\bar{O}$ enables a comparison of the variability of parameters with widely different units and scales, such as ET rate and biomass accumulation, although admittedly, nRMSE fails for $\bar{O}$ = zero or small values close to zero.

2.2.4. k-means clustering

A k-means clustering algorithm was used to group models with similar nRMSE statistics and identify the top-performing models in a non-arbitrary way. Analyses focused on nRMSE for four pairs of model output variables, including simulated ET (ETs) versus grain yield and biomass from −10 to +20 days after planting (DAP) (soil E dominant) and from 41 to 100 DAP (canopy T dominant). Initial tests varied the number of clusters (n) from 1 to 19. The final analysis was conducted with n = 4 clusters based on reducing the sum of squared distance from the cluster center to less than 20% of that for n = 1 cluster. Using n = 4 clusters also resulted visually appealing cluster plots with the set of top-performing models clearly identified within groups having low nRMSE for both variables of each pair. The k-means analysis was conducted using the “scikit-learn” package for Python (Pedregosa et al., 2011).

3. Simulation results and discussion

3.1. Daily results for irrigated and rainfed Mead in 2003 (the driest year) and Bushland 100% and 75% sprinkler irrigations in 2013 (the year with highest observed ET rates)

These four cases were selected from among the twenty treatment-years available for more detailed (daily) examination because they represent the two sites and the two cases at each site with the likely greatest water stress difference between treatments, i.e., irrigated versus

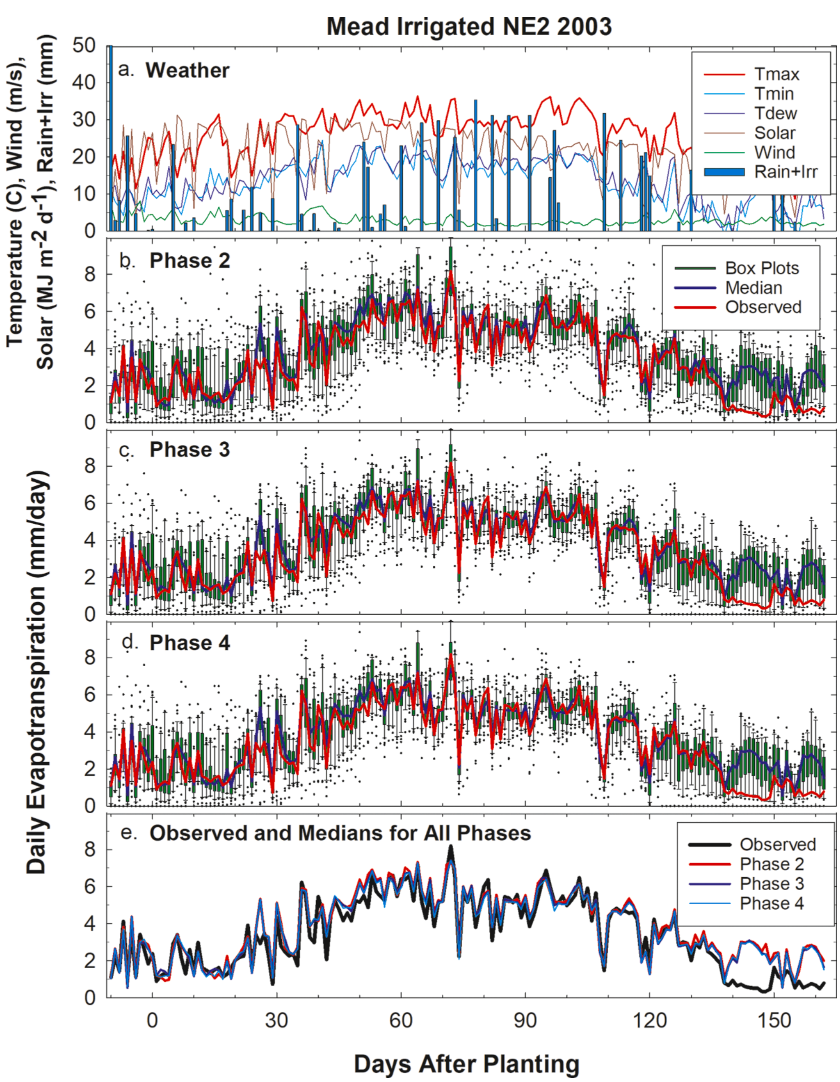

Fig. 1. (a.) Weather variables (maximum and minimum air temperature, dew point, solar radiation, wind speed, rainfall) observed at irrigated field NE2 at Mead in 2003 versus days after planting (DAP). (b.) Box plots of daily simulated evapotranspiration (ETs) where the lower and upper limits of the box indicate the 25th and 75th percentile of ET values simulated by 41 maize growth models, respectively, the lower and upper whiskers indicate the 10th and 90th percentiles, and the points are outliers. Observed values and the median values from the 41 models are also shown. The simulated outputs start with Phase 2, for which the modellers were given leaf area index, growth, and grain yield data for all 100% irrigated treatment-years. Phase 1 simulations, a “blind” test whereby the modellers were only given weather, phenology, management, and soils information, are missing from this graph because a plant population mistake was made for Mead irrigated fields. (c.) Same as (b.) except for Phase 3 whereby the modellers were given the observed ET, soil water content, soil temperature for all 100% irrigated treatment-years. (d.) Same as (c.) except for Phase 4 whereby the modellers were given all data, including ET, growth, and grain yield, for all 20 treatment-years, including rainfed and 75% irrigations. (e.) Observed daily ET values as well as the median ETs values for Phases 2, 3, and 4.

rainfed in the driest year at Mead and 100% versus 75% MESA irrigation in the year with the highest daily ET rates in Bushland.

3.1.1. Daily simulated evapotranspiration (ETs)

3.1.1.1. Irrigated Mead in 2003. As found previously (Kimball et al., 2019), there was a wide range in ETs among the models (Fig. 1). However, the median of all the models tended to be close to the observed values most days. Admittedly, for this intercomparison, as well as for all the others that follow in the rest of this paper, the median is biased toward DSSAT because of the large number of "flavors." For this case, the observed values fell within the short (1–3 mm/d) green boxes much of the time, which indicates many of the models produced respectable simulations. There was only a slight (< 1 mm/d) improvement in model performance going from Phase 1 to Phase 4. It appears that the greatest variability and uncertainty among the models occurred from about 10 days before planting to 10 DAP (soil E dominated) and from about 130 to 160 DAP (after the crop matured). A likely cause of the latter issue is that many models retained a fair amount of green LAI at and after simulated maturity; thus, model equations for ET that depend strongly on LAI did not result in sufficient termination of ET. Successive model adjustments or calibrations going from Phase 1 to Phase 4 as more information was provided only slightly improved this. Model code improvement is needed to decrease green LAI due to senescence, eventually shutting down T at crop maturity. Code improvement likely is also needed to improve the simulation of bare soil ET.

3.1.1.2. Rainfed Mead in 2003. For rainfed conditions at Mead, the models showed large variability (uncertainty) in daily ETs from about −10 to +10 DAP (soil E dominated; Fig. 2) similar to the irrigated field (Fig. 1). The greatest deviations (or errors) occurred from about 70 to 95 DAP when there was little rainfall (Fig. 2a). The observed ET continued at close to 4 mm/d, whereas the models simulated much lower rates. Like the irrigated field (Fig. 1), after DAP 120 as the crop matured, the measured ET decreased rapidly, whereas the models continued to simulate much higher ET. The ET variability during the −10 to +10 DAP

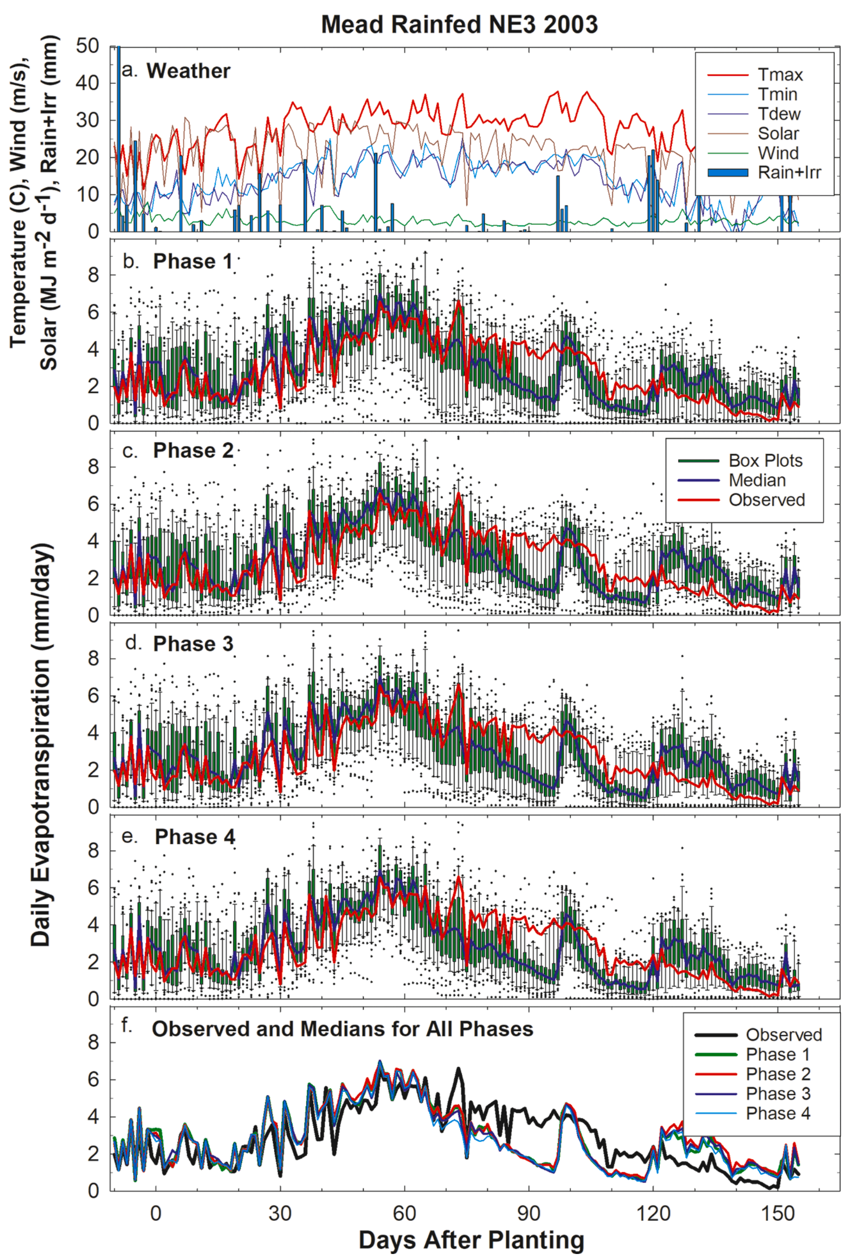

Fig. 2. Similar to Fig. 1 except for rainfed Mead field NE3 in 2003, and data for Phase 1 are also included.

period was related to highly different methods for simulatingE, some of which proved to be less accurate. The issues during the maturation period after 120 DAP are related to the insufficient termination of T after maturity. More importantly, the period from about 70 to 95 DAP and beyond corresponds to the period of water limitation, when most models (and the median) simulated lower than observed ET. We suspect this is caused by inadequate soil water dynamics in the models, such as insufficient rooting depth, inadequate water up-flux or the presence of a perched water table, as well as excessive simulated ET during the early growth phase that depleted the simulated soil water too much, thus reducing ET later.

3.1.1.3. 100% and 75% irrigated Bushland in 2013. For the Bushland location, most of the models (and the median) under-estimated ET during the 45 to 80 DAP period when windspeeds were high (> 5 m/s) and dew points were low (Figs. 3, 4). Model calibration (Phases 1 to 4) only partially improved this situation. This is possibly related to the fact that many of the models do not adequately account for varying wind speed and humidity, as can be deduced from the fact that the models estimated ET fairly well during periods of smaller ET but under-estimated ET greatly during periods of larger ET, when wind speeds were high and relative humidity was low. The fact that solar irradiance was also smaller during some of the periods of smaller ET (due to storm fronts) indicates that the radiation and energy balance algorithms may also need improvement. As before with Mead (Figs. 1, 2), the models failed to reduce T sufficiently after crop maturation (Figs. 3, 4; 105 to 145 DAP). Surprisingly, the models tended to simulate the 75% irrigation treatment (Fig. 4) better than they did the 100% treatment (Fig. 3). Again, we speculate that this is because many of the models had not been calibrated previously to account for the very high winds and low humidity in Bushland, so their ETs simulations were lower than the high observed ET rates for the 100% irrigations treatment (Fig. 3), whereas under the 75% treatment (Fig. 4), drought stress reduced observed ET rates into the ranges for which the models had been calibrated. The fact that observed ET for the 75% MESA irrigations treatment was similar to that for the 100% SDI treatment (Evett et al., 2019) indicates that E may

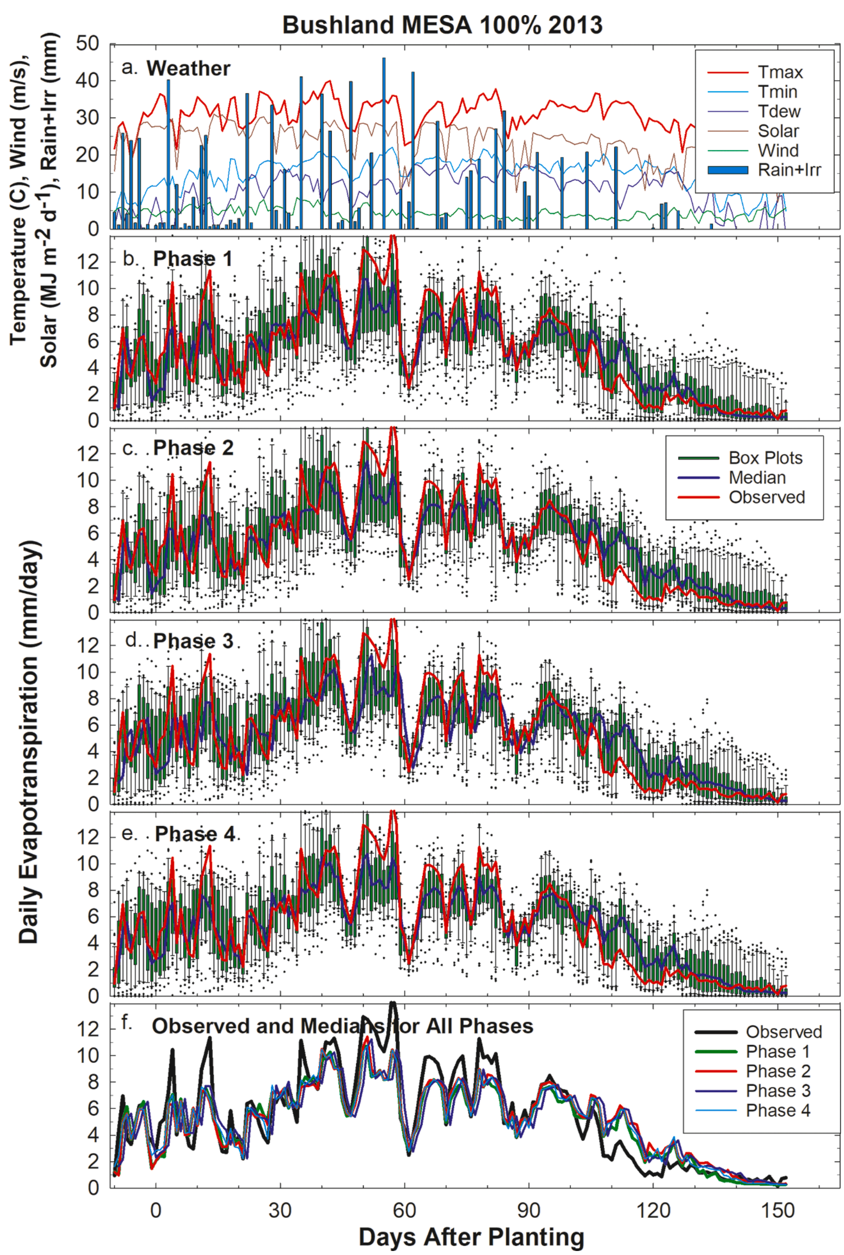

Fig. 3. Similar to Fig. 1 except for 100% MESA (mid-elevation sprinkler application) irrigation at Bushland in 2013.

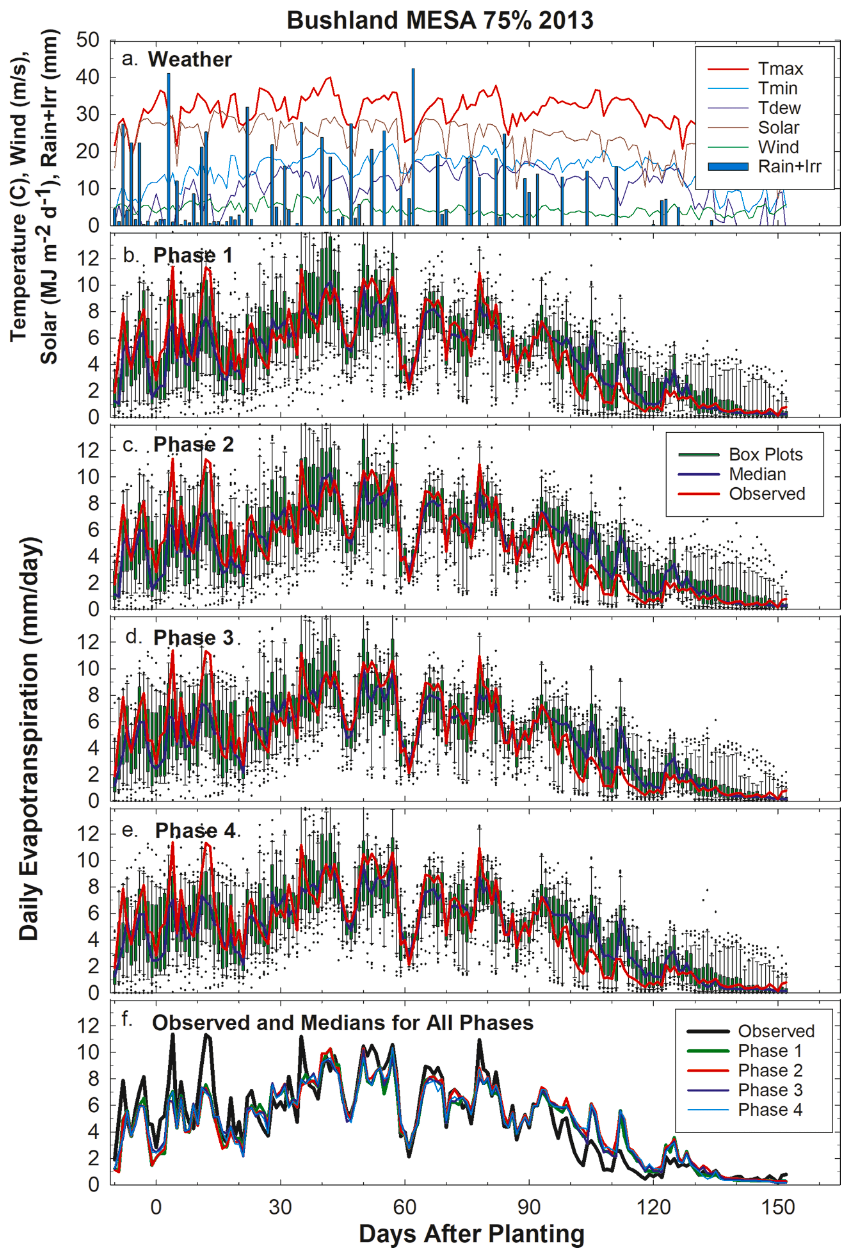

Fig. 4. Similar to Fig. 1 except for 75% irrigation at Bushland in 2013, and data for Phase 1 are also included.

play an important role in the discrepancies between simulated and observed ET for the 100% MESA treatment. The major difference between SDI and MESA irrigation in the Bushland experiments was the larger evaporative losses from the soil surface in MESA irrigated fields (Evett et al., 2019).

3.1.2. Ranking of models with respect to their nRMSE for simulating daily ETs

3.1.2.1. Irrigated mead in 2003. The median of all the models had the lowest nRMSE for ETs for Phases 2, 3, and 4 for both early season (−10 to +20 DAP; soil E dominant) and mid-season (41 to 100 DAP; canopy T dominate) (Fig. 5). For early season STCK was the best model followed by several DSSAT “flavors,” and at mid-season several DSSAT flavors again did well, especially for Phase 2. STCK uses Penman (1948) to calculate atmospheric demand and the 2-phase model of Brisson and Perrier (1991) and Brisson et al. (1998, 2003) to calculate soil water evaporation, Ea (Table S1). Note that all the DSSAT flavors listed for −10 to +20 DAP end in “R”, which indicates that the soil E method of Ritchie (1972) was better than the more recent method of Suleiman and Ritchie (2003, 2004). However, for Phase 2 during the 41 to 100 DAP period DIFS and DCFS did well, but during this period canopy T was dominant, so soil E was relatively unimportant then.

The effects of changes made by the modelers going from phase to phase can also be seen in Fig. 5. For example, BIOM was ranked 19th for Phase 2, −10 to +20 DAP but improved to 5th and 6th for Phase 3 and Phase 4, respectively. Like the well-performing DSSAT flavors, BIOM also uses Ritchie (1972) to simulate soil E. AHC rose from 28th to 5th from Phase 2 to Phase 4 for the −10 to +20 period. AHC uses the two-stage FAO-56 method to simulate E for mostly bare soil (Table S1). A huge improvement was made by SLUS going from 40th for Phase 2 to 5th for Phase 3 for the 41 to 100 DAP period. SLUS calculates atmospheric demand from Priestly and Taylor (1972) and then uses an empirical equation to simulate potential ETp (Table S1), which would be mostly T for the irrigated full canopy. XNSM, SMET, and CS all markedly improved from Phase 2 to Phase 4 to be among the best for the full

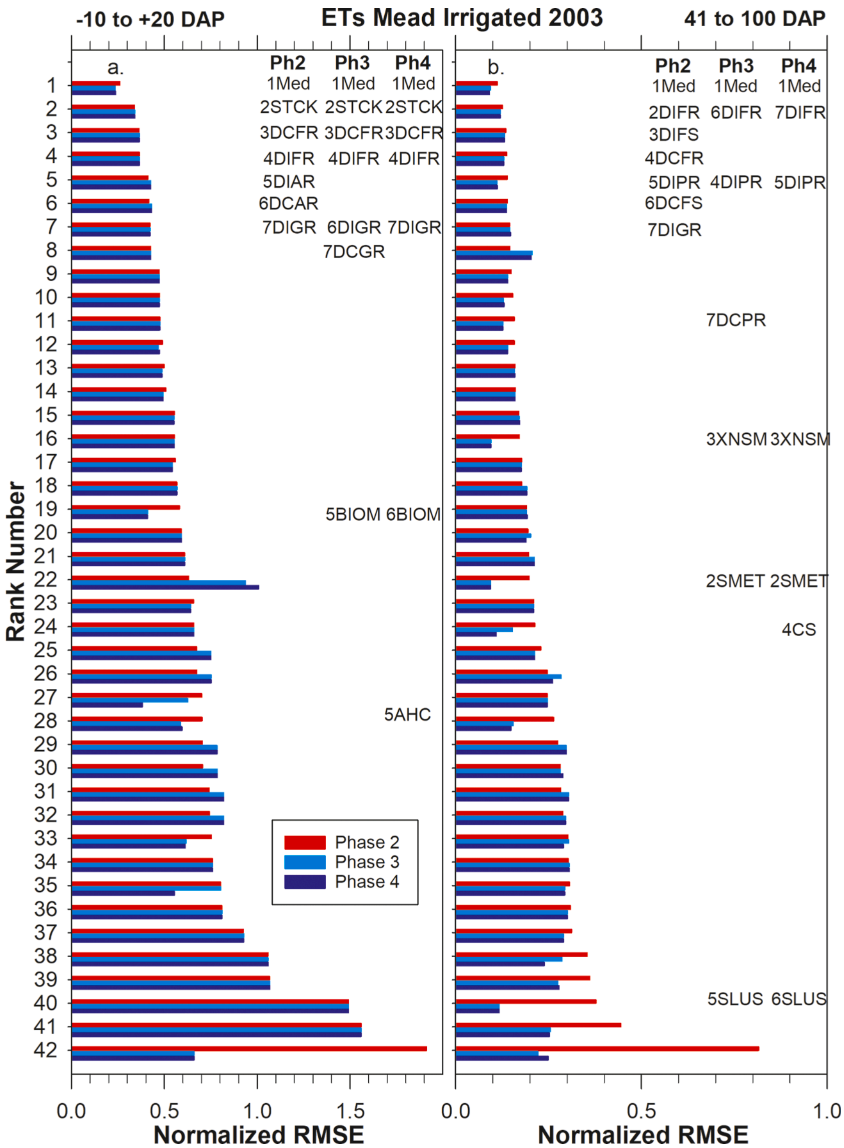

Fig. 5. (a) Normalized root mean squared error (nRMSE) between observed and simulated daily ET values from −10 to +20 days after planting (DAP)(mostly soil E) for the irrigated field NE2 at Mead in 2003 for all the models. Phases 2, 3, and 4 are identified by red, cyan, and blue bars with Phase 2 at the top and Phase 4 at the bottom of each group. Phase 1 data are missing from this graph because a plant population mistake was made for Mead irrigated fields. The models have been sorted in ascending order of nRMSE for Phase 2 from top to bottom of the graph with the rank numbers on the left axis indicating their ranking for Phase 2. The Median (Med) and the six best models (lowest nRMSE) for Phase 2 are listed under "Ph2". Somewhat similarly, the Median and six best models for Phases 3 and 4 are also listed under "Ph3" and "Ph4", but because the modelers made different adjustments going from phase to phase, their rank order changed, so the names along with their nRMSE rank are in different positions down the graph. (b) Same as for (a) except the data are for 41 to 100 DAP (mostly crop canopy T) with the ranking done on the 41 to 100 DAP Phase 2 data.

canopy (Fig. 5b). All three use FAO-56 (Allen et al., 1998) with some modifications (Table S1).

3.1.2.2. Rainfed mead in 2003. The STCK model was best for simulating ETs for the −10 to +20 DAP period in the rainfed field at Mead in 2003 for Phases 1 and 2, while the median was 2nd, and then they traded rankings for Phases 3 and 4 (Fig. 6a). ECOS, JUL, DCFR, DIFR, and STSW also did very well. ECOS is a full energy balance model while JUL uses the Penman-Monteith approach (Monteith, 1965) with a 10-layer canopy (Table S1). BIOM rose from 31st for Phase 1 to 3rd for Phases 3 and 4. JUL and XNGM were best for the 41 to 100 DAP period (Fig. 6b). MZD was 3rd for Phase 1, but did much worse in the other phases. DIFR, DCFR, and DIPR did well. ECOS rose from 13th for Phase 1 to 2nd for Phases 3 and 4. All of these listed models were better than the median for this case.

3.1.2.3. 100% MESA irrigation at Bushland in 2013. The median of all the models ranked 1st at simulating ETs from −10 to +20 DAP for all phases in Bushland with 100% MESA irrigation in 2013 (Fig. 7a). Except for Phase 2, ECOS, an energy balance model was best. DCFR, DCPR, STSW, DIFR, and XNGM all did well. At mid-season (41 to 100 DAP, Fig. 7b), CS, DIGR, and STCK did well for all phases. BIOM and DIGS improved greatly for phases 3–4. However, the median was only about 12th.

3.1.2.4. 75% MESA irrigation at Bushland in 2013. The median of all the models was 1st for all but one phase for both early season (−10 to +20 DAP) and mid-season (41 to 100 DAP) for the 75% irrigation treatment at Bushland in 2013 (Fig. 8). MZH was ranked 2nd for Phase 1, early season (Fig. 8a) but then did much worse for other phases. Similarly, ARMO did well for Phases 1 and 2, but then did much worse. DCFR, CS, and DIPS did well in all phases. XNGM, AMSW, and DCPS were among the best for Phase 4. AMSW uses a transpiration efficiency to compute ETs from biomass accumulation, XNGM uses a modified Penman-Monteith Monteith (1965), and DCPS uses Priestly and Taylor (1972) to simulate potential atmospheric demand and ultimately ETs.

At mid-season, AQCP and AMSW did well for all phases (Fig. 8b). MZH and MZD did well for Phase 1, but then much worse for later phases. DIFS, DCFS, SMET, and BIOM did well for Phases 3 and 4.

3.1.2.5. Intercomparison among the models for all four cases of daily et for Phase 4. Looking at Figs. 5-8, no single model appears among the best (lowest nRMSE) six for all four cases. The median was among the best for the all four cases from −10 to +20 DAP (mostly E), but only for two cases

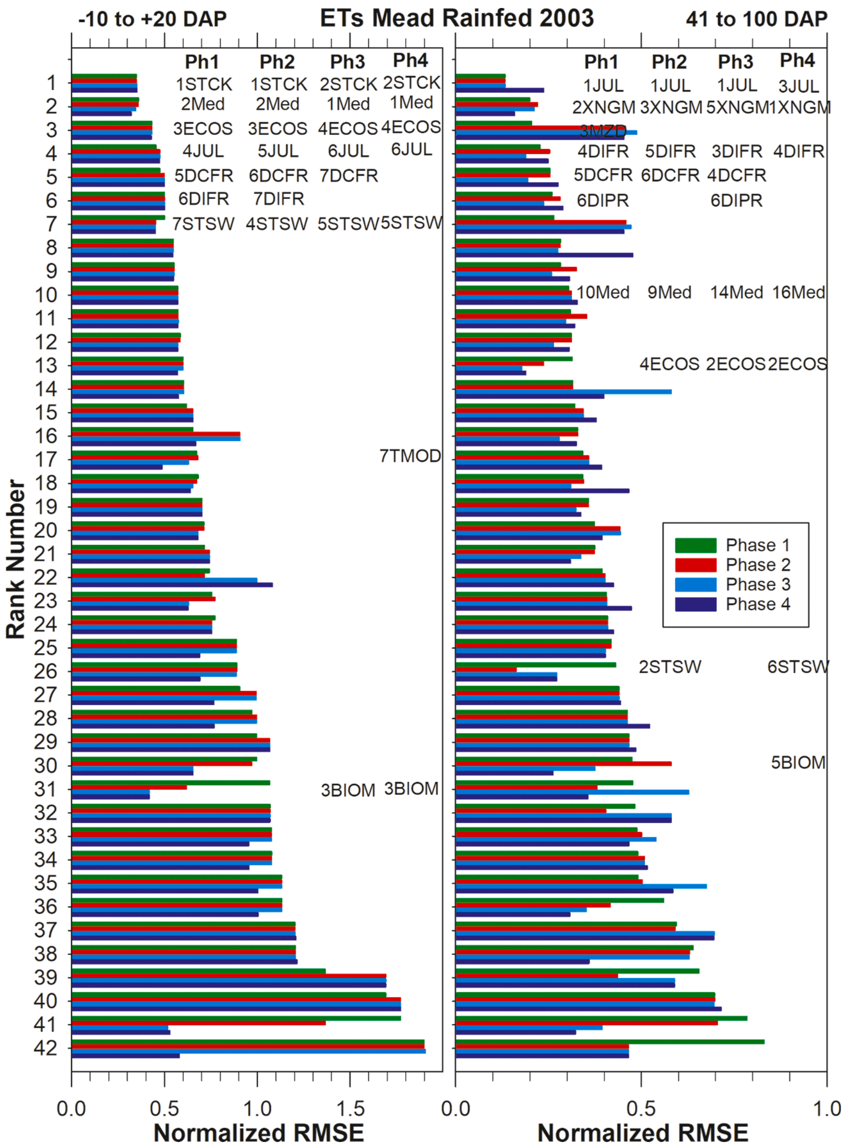

Fig. 6. (a.) Normalized root mean squared error (nRMSE) between observed and simulated daily ET values from −10 to +20 days after planting (DAP)(mostly soil E) for the rainfed field NE3 at Mead in 2003 for all the models. Phases 1, 2, 3, and 4 are identified by green, red, cyan, and blue bars with Phase 1 at the top and Phase 4 at the bottom of each group. The models have been sorted in ascending order of nRMSE for Phase 1 from top to bottom of the graph with the rank numbers on the left axis indicating their ranking for Phase 1. The Median (Med) and the six best models (lowest nRMSE) for Phase 1 are listed under "Ph1". Somewhat similarly, the Median and six best models for Phases 2, 3 and 4 are also listed under "Ph2", "Ph3", and "Ph4", but because the modelers made different adjustments going from phase to phase, their rank order changed, so the names along with their nRMSE rank are in different positions down the graph. (b.) Same as for (a.) except the data are for 41 to 100 DAP (mostly crop canopy T).

from 41 to 100 DAP (mostly T). Focusing on the −10 to +20 periods (mostly E), DCFR was among the best for 3 cases; STCK, DIFR, BIOM, ECOS, STSW, SNGM, and AMSW for 2 cases; and DIGR, JUL, TMOD, CS, DIPS, DCPS for 1 case. For the 41 to 100 DAP periods, the median was among the best only twice. BIOM was best for 3 cases; DIFR, CS, and SMET were best for 2 cases; and STCK DIGR, ECOS, JUL, STSW, XNGM, DIPR, XNSM, SLUS, DIGS, SLFT, AQCP, AMSW, DIFS, and DCFS were all among the best for 1 case. BIOM stands out as being the only model to be among the best twice for early season (mostly E) and thrice for mid-season (mostly T).

3.2. *Inter-comparisons within the DSSAT family*

3.2.1. *Daily ETs*

A comparison of E methods within the DSSAT models, revealed that the older Ritchie-2-stage model (Ritchie, 1972) was consistently better (lower nRMSE and lower simulated ETs) than the Sulieman and Ritchie method (2003, 2004) during the −10 to +20 DAP period, regardless of the other ET methods (Figs. 9a, 10a). The Ritchie-2-stage method was also better (slightly lower nRMSE) for ETs in the 41 to 100 DAP full canopy phase (Figs. 9b, 10b) for two reasons (less E during that phase, but mostly because lower early E allowed soil water in deeper layers to be conserved for the 41 to 100 DAP period, thus contributing more to T during the latter phase).

In spite of having a theoretically more realistic mechanism for moving soil water with potential gradients, the Hydrus method (ŠŚŠimůnek et al., 1998, 2008; Shelia et al., 2018) did not perform as well as the more empirical Ritchie (1972) and Sulieman and Ritchie (2003, 2004) methods (Fig. 9a, 9b). However, Hydrus was just recently incorporated into the DSSAT shell, whereas the Ritchie (1972) and the Sulieman and Ritchie (2003, 2004) routines have been used for many years and likely have been fine-tuned to the system. Also, Hydrus is very sensitive to the values of the soil physical and hydraulic properties, so if those parameter values were off, the simulated ET would also be off.

A comparison of potential ET (ETp) methods within the DSSAT models illustrated that the FAO-56 method (in present DSSAT; Allen et al., 1998) with Kcan of 0.62 (gives Kep = 0.50) performed better (lower nRMSE) for ETs than the other ETp methods: Priestley-Taylor (P-T; 1972), alfalfa reference-[ET_r, ASCE *equation* (R.G. Allen et al., 2005)], or grass reference-[ET_o, ASCE equation (R.G. Allen et al., 2005)] during both the −10 to +20 DAP period and the 41–100 DAP period (Figs. 10a, 10b). Kcan is the extinction coefficient for absorption of photosynthetically-active radiation by LAI, while Kep is the extinction coefficient for absorption of total solar energy by LAI. The default Kcan

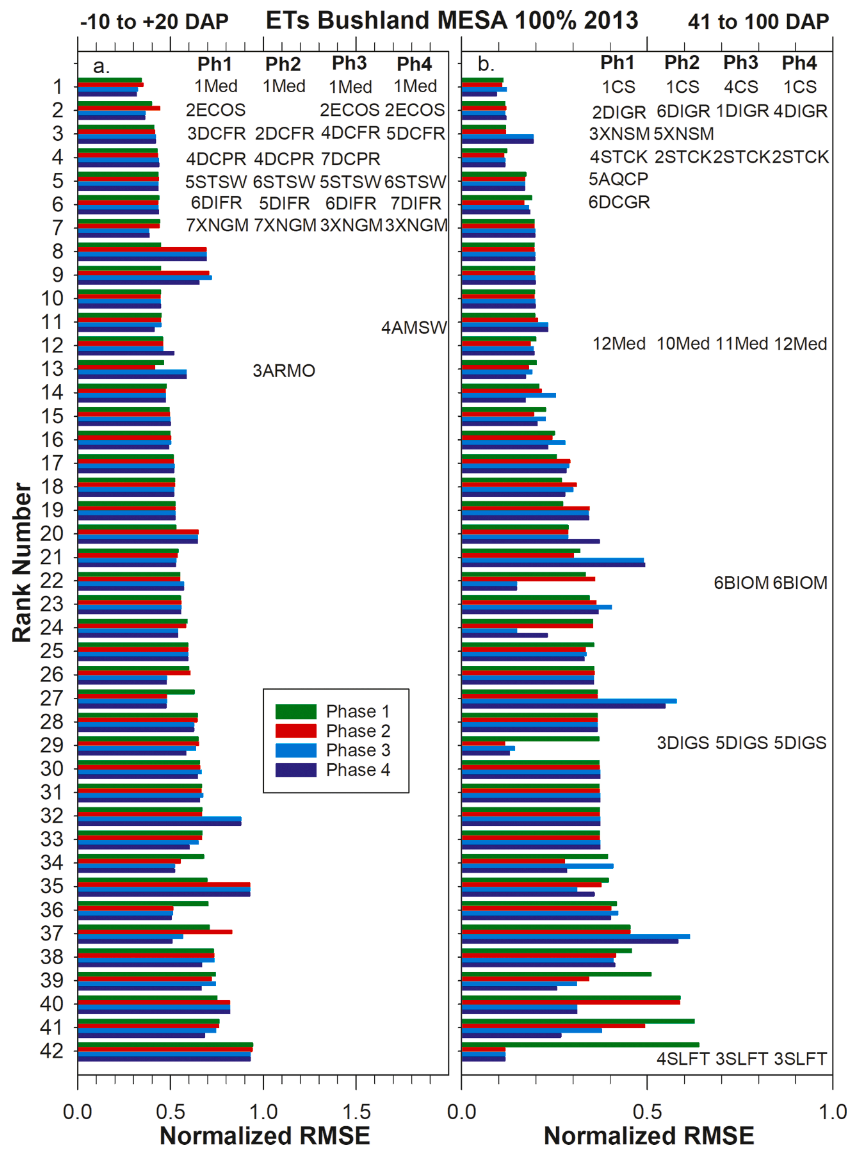

Fig. 7. Like Fig. 6 except for 100% MESA (mid-elevation sprinkler application) irrigation to restore soil water to field capacity at Bushland.

for CERES is 0.85 (in the ecotype file). Kcan was reduced to 0.62 during phase 3, which reduces the effective energy extinction from 0.685 to 0.50 [latter value supported lysimeter studies of Villalobos and Fereres (1990), as well as the theory of foliar absorption of total solar energy (Goudriaan, 1977)]. The Kep=0.50 was used for P-T as well. On the other hand, the alfalfa reference-FAO-56, or grass reference-FAO-56 are dual-coefficient methods that compute their own coefficients during incomplete and full canopy phases of ET. In contrast to a previous study on cotton (*Gossypium hirsutum* L.) ET (Thorp et al., 2020), the methods based on ASCE alfalfa and grass reference ET did not perform as well as DSSAT FAO-56 and P-T; however, the calibration methodology limited their comparability in the present study. It appears that the newly reduced Kep of 0.50 contributed to improved DSSAT performance, and it is an improvement over the default DSSAT value. As mentioned previously, Sau et al. (2004) reported that the FAO-56 with a Kep=0.50 gave the best simulations of ET, soil water extraction, and biomass accumulation with the CROPGRO-Faba bean model for a water-limited environment. FAO-56 was better than P-T, and the extinction coefficient (Kep=0.50) was better than a higher Kep for either ET method. Similarly, Lopez-Cedron et al. (2008) found the CERES model gave better simulations of maize biomass, grain yield, and harvest index under water-limited environments, using FAO-56 rather than P-T, and again, Kep of 0.50 was better than a higher energy extinction coefficient (default in CERES was 0,685).

There was no significant difference in nRMSE between the CERES or IXIM models for the −10 to +20 DAP period (Fig. 10a, soil E dominant), whereas for the 41 to 100 DAP period (Fig. 10b, canopy T dominated), IXIM was slightly better, likely because of its more realistic simulation of LAI progression. IXIM senesces green leaf area more rapidly (and more mechanistically) near maturity than does CERES, which results in less T during the grain-filling phase, and which more closely matches the observed reduction inT.

Comparing methods for calculating potential evapotranspiration (ETp) on the nRMSE of ETs for the 41 to 100 days after planting (DAP) period (Fig. 10b), the FAO-56 method had significantly lower nRMSE. For the −10 to +20 DAP period (Fig. 10a), it was better than both the alfalfa (tall; *Medicago sativa* L.) and grass (short) crop coefficients with the ASCE standardized reference equation, but Priestley-Taylor (P-T) tended to be almost as good. Comparing soil E methods, Ritchie (1972) was much better than Suleiman and Ritchie (2003, 2004) for the −10 to +20 DAP period (Fig. 10a) soil E dominant), and Ritchie (1972) was slightly better even for the 41 to 100 DAP period (Fig. 10b, canopy T dominant).

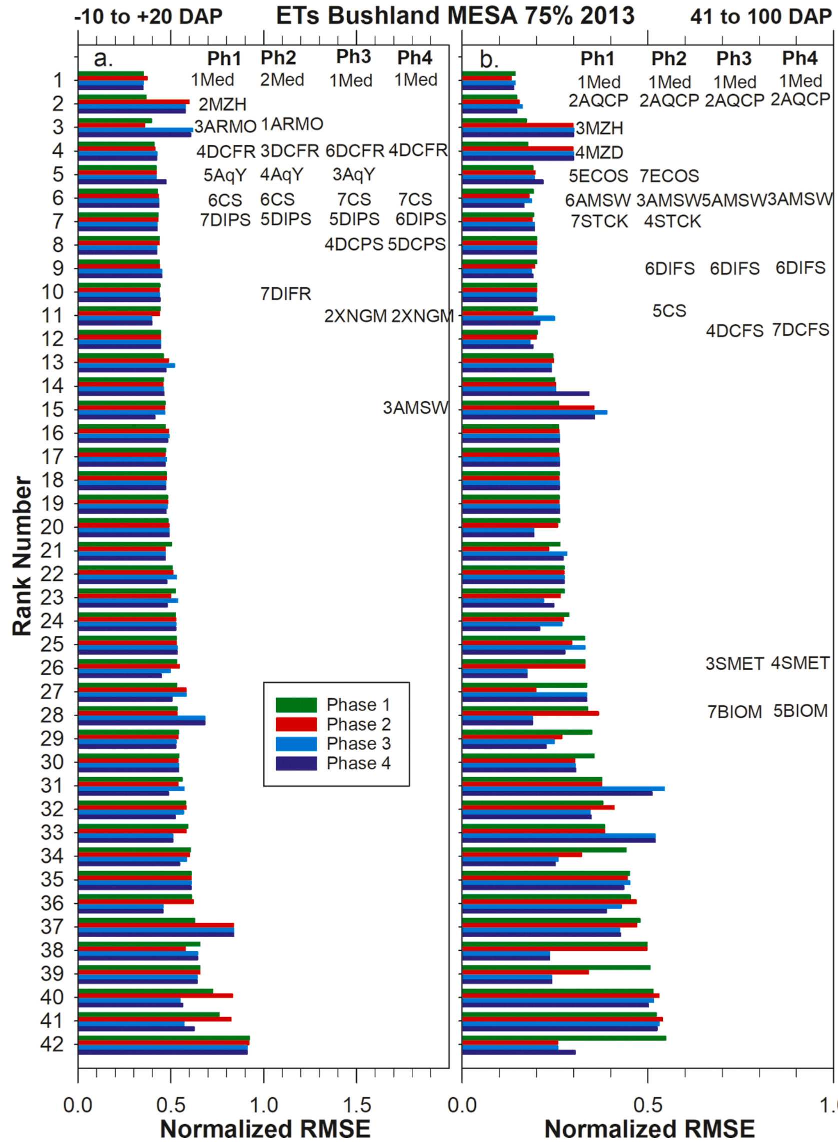

Fig. 8. Like Fig. 6 except for the MESA (mid-elevation sprinkler application) 75% irrigation at Bushland.

3.3. *Inter-comparisons within the STICS, Expert-N, and MAIZSIM families*

Comparing the nRMSE of daily ETs between the two "flavors" of each pair of the STICS, Expert-N, and MAIZSIM families, there were no significant differences (Figs. 11a, 11b). STCK uses a single surface model (Penman, 1948) to compute potential ETp, whereas STSW handles separate canopy and soil surfaces (Shuttleworth and Wallace, 1985), Thus, for these four cases, STCK and STSW performed equally well at simulating soil E (Fig. 11a) and canopy T (Fig. 11b) in spite of the different methods for simulating ETp. Both XNSM and XNGM models use Penman-Monteith based approaches for simulating ETp. However, XNSM follows FAO 56 guideline based on ETo multiplied with a single crop factor to get ETp while in XNGM the required surface- and aerodynamic resistances are calculated directly from simulated LAI and simulated canopy height. In addition, XNGM follows the more detailed Farquhar model in simulating photosynthesis and leaf T but simplifies vertical root distribution. The latter could possibly explain slightly better soil moisture simulations of XNSM compared to XNGM (data not shown). In XNSM, temperature, moisture, and nutrient availability in different soil layers are taken into account when simulating rooting depth and root length distribution. In contrast, XNGM assumes a uniform distribution of root length density within the rooted zone, with the increase in rooting depth simply simulated from the increase in root biomass, regardless of the soil conditions. Thus, considering that there are marked differences between the two models, it is surprising that they differ so little in their ability to simulate ETs. The lack of significant differences between MZD and MZH is reasonable because they are the same in their representation of plant and soil processes. Both models run on an hourly time step internally but MZD takes daily weather data as input and interpolates them into hourly time steps, while MZH takes hourly weather data directly as input.

3.4. *Potential ETp and other sources of variability/error in daily ETs*

There was a wide variability among the models in their simulated values for daily ETs as shown in Figs. 1-4, which is similar to the previous results reported by Kimball et al. (2019). In that report, Fig. 10 shows that much of the variability can be attributed to variability among the models in their values of ETp. Therefore, for this study we requested more values of "upstream" variables that the modelers might be using to compute ETp, including reference ET based on short (12 cm) grass (ETo), reference ET based on tall (50 cm) alfalfa (ETr), soil coefficient (Ks), basal crop coefficient (Kcb), soil evaporation coefficient for drying

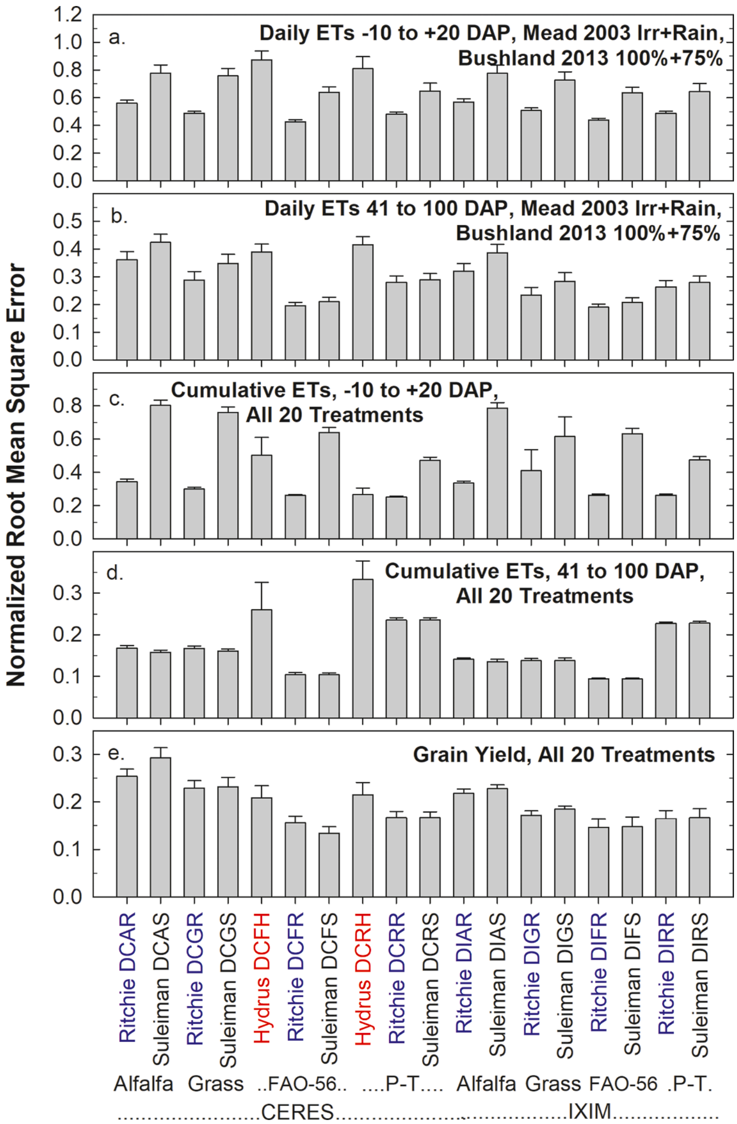

Fig. 9. Normalized root mean square errors (nRMSE) of the 18 “flavors” of the DSSAT family models (a) for the −10 to +20 DAP periods (mostly soil E) of daily ETs over all phases for the irrigated and rainfed data for Mead 2003 and the 100% and 75% MESA irrigated data for Bushland 2013. Models included are DSSAT CSM-CERES and DSSAT CSM-IXIM, whose horizontal names span the corresponding left ten and right eight vertical bars, respectively. Potential ETp calculation methods are using alfalfa (tall, ET_r) and grass (short, ET_o) reference crop coefficients with the ASCE standardized reference equation (R.G. Allen et al., 2005), FAO-56 (Allen et al., 1998), and Priestley-Taylor (1972). These horizonal names span the corresponding bars above them. Soil evaporation calculation methods follow Ritchie (1972; labelled “Ritchie”), Suleiman and Ritchie (2003, 2004; labelled just “Suleiman”), and Hydrus (Šimůnek et al., 1998, 2008; Shelia et al., 2018; labelled “Hydrus”). (b) Like (a) except for the 41 to 100 DAP periods. (c.) The values plotted are averages (+ standard errors) of the nRMSEs for Phase 4 for all 20 treatment-years of the cumulative ETs from −10 to +20 DAP periods. (d.) Like (c.) except for the cumulative ETs from 41 to 100 DAP. (e.) nRMSEs for Phase 4 grain yields for all 20 treatments.

soil (Ke), overall crop coefficient (Kc), potential soil evaporation (Ep), potential transpiration (Tp), ETp, and of course, ETs. Three of the models did not report ETp, presumably the energy balance ones that do not use the concept.

Focusing on the Phase 2 results from irrigated Mead in 2003, 34 models reported Ep and 35 models reported Tp, and both Ep and Tp were quite variable (data not shown), As expected, the magnitude and variability of the soil Ep were greatest for bare soil at the beginning of the season. However, there was more than a 2 mm/day spread even at mid-season. Surprisingly, a few of the models showed some Tp starting on the day of planting before the plants had even emerged. Then, as the Tp increased in magnitude as plants grew to full size by mid-season, so did the range in variability among them, similar to ETs.

Thirteen of the models reported ETo and only 4 reported ETr. Presumably ETo and ETr depend only on weather, yet ETo varied by a factor of about 2 at midseason among the 13 models (data not shown). Apparently, several different definitions and equations for ETo are in play among these models.

Only 6, 4, 4, and 7 models used Ks, Kcb, Ke, and Kc, respectively. It seems likely that more models do use them, but they are computed and not routine output, so the modelers would have had to change code to get them. In any event, there appear to be several ways that models are getting from ETo (or ETr) to ETp that are contributing to the variability of ETs.

Thus, in conclusion, the variability in ETp and ETs appears to be coming from steps all along the way starting from the calculations of ETp to the final resultant ETs.

3.5. Cumulative month to whole season ETs results for all 20 treatment-years

The previous sections focused on the daily ET for four selected treatment-years. However, one can imagine that an underestimate of simulated daily ET one day could save some simulated soil moisture and lead to an overestimate the next day. The following sections examine the cumulative ET over longer time periods to reveal the extent that the errors are also cumulative.

3.5.1. ETs from −10 to +20 DAP (mostly soil E) and 41 to 100 DAP (mostly canopy T)

Moving from daily ETs for the four cases (2003 for irrigated and rainfed Mead; 2013 MESA at 75% and 100% irrigation at Bushland) to cumulative ETs over longer time durations for all 20 treatment-years also showed wide variability among the models (Fig. 12). Again, there were variations by factors of 2 to more than 4 among them in cumulative ETs from −10 to +20 DAP (mostly soil E) (Fig. 12a). There was little or no improvement in going from Phase 1 to Phase 4. For Treatments 1–10

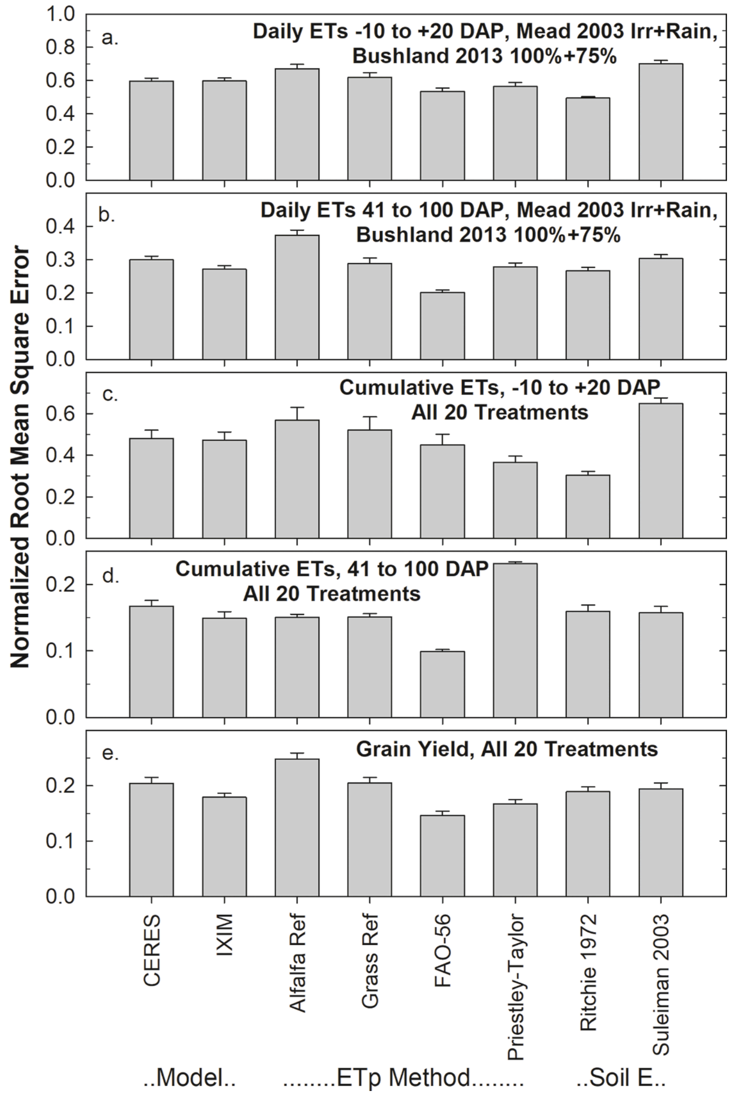

Fig. 10. Direct comparisons using the same data as for Fig. 9 (excluding Hydrus) between the DSSAT-CERES and DSSAT-IXIM models, among the four potential ET methods, and between the two soil water evaporation methods for the corresponding a, b, c, d, and e graphs. The horizontal "Model", "ETp Method", and "Soil E" labels span the corresponding bars above.

for Mead, the medians of the models were close to the observations, but for Treatments 11 and 12, the models generally overestimated ETs. For Bushland, most of the models underestimated Treatment 13 when spray irrigation wetted the surface and Treatment 18 when rainfall wetted the surface of SDI fields. Most models overestimated Treatments 17 and 19 when the SDI field surface was dry despite plentiful irrigation, but the medians were close to observed for the other 4 treatments. These results indicate problems simulating E from wetted surfaces and with simulated redistribution of water from buried drip lines to the surface (too much water movement to the surface).

Looking at cumulative ETs from 41 to 100 DAP (mostly canopy T), there is a range of about a factor of 2 among the models (Fig. 12b), which is bad but less than that from the bare soil (Fig. 12a). For Mead, most of the models overestimated ETs for Treatments 1–6 and 8–11. They underestimated Treatment 7 but were close for Treatment 12. For Bushland, most of the models underestimated ETs under sprinkler irrigation for Treatments 14–16, which represent wetter soil Ranking the models' ability to simulate cumulative ETs from −10 to +20 DAP by nRMSE (Fig. 13a), the medians were close to observations for Phases 2–4. SLFT was the best model for Phases 2 and 3 and was next best in Phase 4. SLFT uses FAO-56 (Allen et al., 1998) to calculate atmospheric demand and then dual crop coefficients simulate ETs (Table S1). For Phase 2, models in the DSSAT family were ranked 3–7, and several did well in Phases 3 and 4. AMSW was best in Phase 4. CS and XNGM were among the best in Phases 3 and 4.

Similarly ranking their ability to simulate ETs from 41 to 100 DAP (Fig. 13b), several of the models in the DSSAT family did well for Phases 2, 3, and 4. ECOS was among the best for Phases 2 and 3. SSMi (which uses Priestly and Taylor (1972) for potential atmospheric demand and transpiration efficiency with biomass accumulation to simulate ETs) was ranked 6 for Phases 3 and 4, and SMET was ranked 3rd for Phase 4. surface conditions, but the medians were close for Treatment 13. Under SDI irrigation, most models underestimated Treatment 18, but the medians were close for Treatments 17, 19, and 20.

Looking back at Section 3.1.2.5, BIOM was among the best at simulating daily ETs, yet it was not among the best at simulating ETs over the longer intervals. On the other hand, DIFR was almost as good as BIOM for simulating daily ETs, and it was best for simulating cumulative ETs over the 41 to 100 DAP periods (Fig. 13, Phase 4). Besides DIFR, DCFS and SMET were the only other two models that were among the best for cumulative ETs over the 41 to 100 DAP periods and also were among the best for at least one case of daily ETs. For the −10 to +20 DAP periods, DCFR, XNGM and CS are the only models that were best for simulating cumulative ETs and also for daily ETs at least for one case.

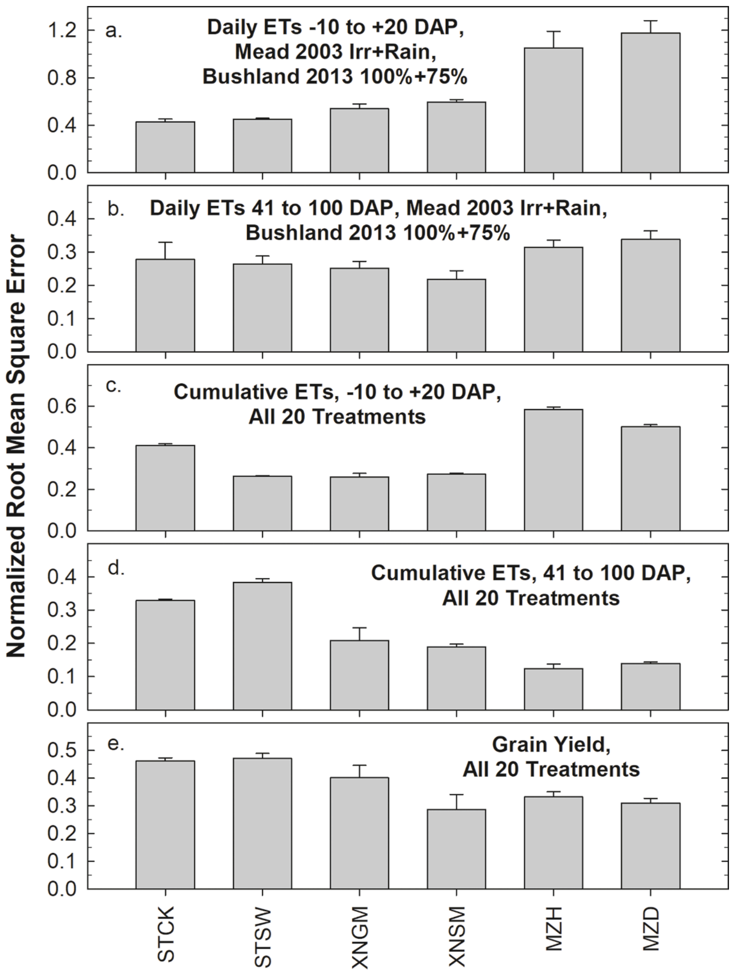

Fig. 11. Direct comparisons using nRMSE between the STCK and STSW flavors of the STICS model family, between XNGM and XNSM flavors of Expert-N family, and between the MZH and MZD flavors of the MAIZSIM model for (a) the −10 to +20 DAP time period (mostly soil E). The data used were all phases for the irrigated and rainfed data for Mead 2003 and the 100% and 75% MESA irrigated data for Bushland 2013. (b). Like (a) but for the 41 to 100 DAP period (mostly canopy T). (c.) Phase 4 of cumulative ETs from −10 to +20 DAP for all 20 treatments. (d) Like (c) but for cumulative ETs from 41 to 100 DAP. (e) Phase 4 grain yield for all 20 treatment-years.

Thus, doing well for simulating daily ETs did not guarantee success at simulating cumulative ETs.

3.5.2. Inter-comparisons of cumulative ETs within the DSSAT and other model families

There were wide differences in performance among model "flavors" within the DSSAT family for cumulative simulated ETs from −10 to +20 DAP (mostly soil E) over the 20 treatment-years (Fig. 9c). Most obvious is that the Ritchie (1972) soil E method did much better than the corresponding Suleiman (Suleiman and Ritchie, 2003, 2004) method for every case. The Hydrus method did comparatively well for this cumulative-ETs/20-treatment-year comparison, which is in contrast to the daily-ETs/4-treatment-year comparison in Fig. 9a.

As would be expected, looking at the 41 to 100 DAP periods, the soil E method had little effect (Fig. 9d). However, Hydrus, did poorly which is in contrast to the −10 to +20 periods (Fig. 9c).

There was no significant difference in performance between the CERES Maize and IXIM Maize models for the −10 to +20 DAP periods (Fig. 10c), whereas IXIM was slightly better than CERES for the 41 to 100 DAP periods (Fig. 10d). The better performance of IXIM for full canopy conditions was likely because of its more realistic simulation of LAI progression, as mentioned previously. Priestley-Taylor was the best ETp method for the −10 to +20 DAP periods (Fig. 10c) but worst for the 41 to 100 DAP periods (Fig. 10d). FAO-56 was second best for −10 to +20 DAP periods (Fig. 10c) but best for the 41 to 100 DAP periods (Fig. 10d). As was obvious from Fig. 9c, the direct comparison between Ritchie (1972) and Suleiman and Ritchie (2003) in Fig. 10c, confirms the superiority of the older Ritchie (1972) method for simulating soil E, likely because the Suleiman and Ritchie overestimates the upward movement of soil water from deeper depths. However, under full canopy conditions (Fig. 10d), there was no difference between the two soil E methods.

Looking at other models with more than one flavor, STSW performed better than STCK for cumulative ETs from −10 to +20 DAP (Fig. 11c), but the reverse was true from 41 to 100 DAP (Fig. 11d). It is somewhat surprising that the two-surface method for computing ETp in STSW did better for the −10 to +20 DAP period when there was only the single soil surface, but was worse for the 41 to 100 DAP full canopy period. However, looking more closely, both models did well for both time periods for the Mead data, whereas for Bushland in 2013, both models had trouble getting emergence with SDI in 2013, and this issue distorted the results. There was no significant difference between XNGM and XNSM for either of the time periods (Figs. 11c, 11d). As noted in Section 3.3, the two models use slightly different variants of the Penman-Monteith approach and differing root distribution approaches resulting in essentially no differences in daily ETs for the four cases (Figs 11a, 11b) nor in cumulative ETs for all 20 treatments (Figs. 11c, 11d). MZD did slightly better than MZH for the −10 to +20 DAP periods (Fig. 11c), but there was little difference for 41 to 100 DAP (Fig. 11d). Any differences between MZD and MZH are likely associated with the differences between interpolated and measured hourly weather data that were driving MZD and MZH, respectively.

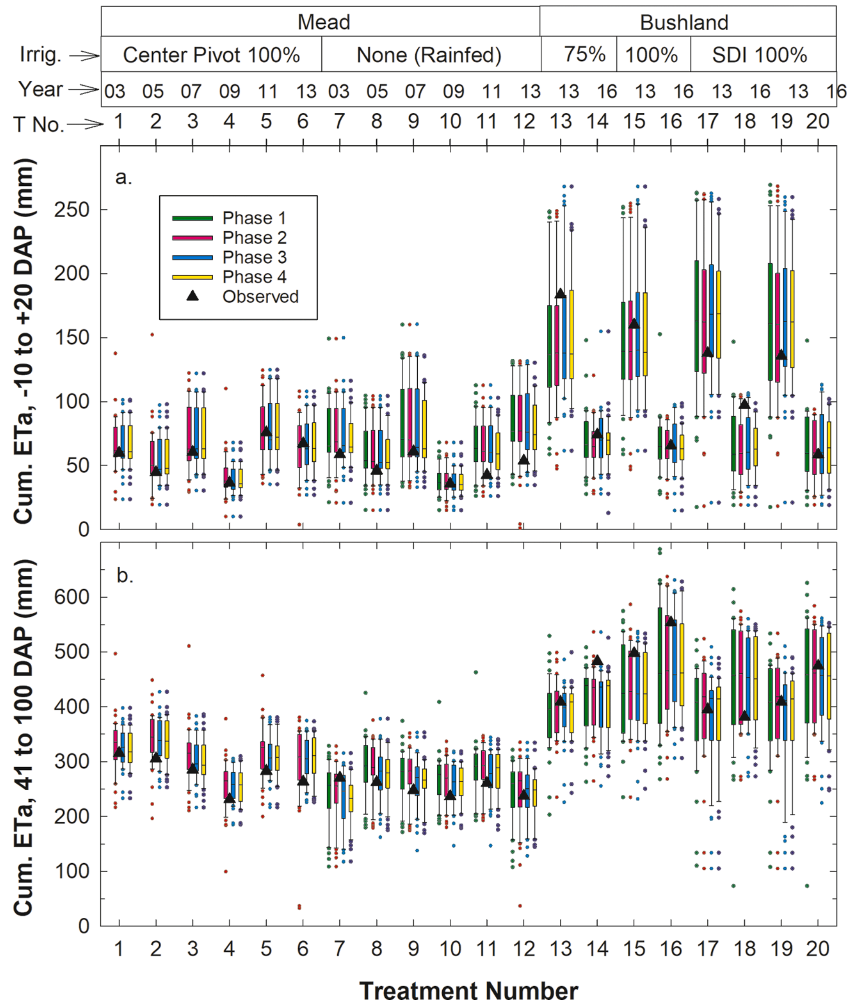

Fig. 12. Box plots for all 20 treatment-years (as defined at the top) of cumulative simulated evapotranspiration (ETs) over (a) the −10 to +20 days after planting (DAP) time period (mostly E) and (b) the 41 to 100 DAP time period (mostly T) for all four phases. The dark lines across the boxes indicate the medians of all the models. Also shown are the corresponding observations. Phase 1 is not shown for treatments 1–6 because of a planting density mistake.

3.6. Ability of the models in Phase 4 to simulate agronomic parameters for all 20 treatment-years – maximum leaf area index, biomass at about 40 DAP and about 100 DAP, and final grain yield

3.6.1. Considering all the models

There was a wide range in simulations of maximum LAI between the lowest and the highest models (Fig. 14a). However, for some treatments, most of the models agreed closely as indicated by short boxes. Indeed, for Treatments 1 and 3, most of the models agreed almost exactly with one another and with observations. For many treatments, the medians agreed closely with observations. However, for Treatments 4 and 14, the models mostly underestimated LAI, whereas for Treatments 11, 15, and 17, they mostly overestimated LAI. For Bushland, treatments 15, 17, and 19 were in the 2013 year that began quite dry and required plentiful irrigation to achieve germination and to support crop growth. Overestimation of LAI may be linked to model algorithms overreacting to the plentiful irrigation in an otherwise stressful year.

Most of the models overestimated above-ground biomass at about 40 DAP for almost all the treatments (Fig. 14b). This was particularly true for the dry 2013 year at Bushland, again indicating that the plentiful irrigation caused the models to overestimate biomass accumulation despite an otherwise stressful environment. However, by 100 DAP (Fig. 14c), most of the models did much better, and agreement with observations was much closer. For final grain yield, most of the models did surprisingly well (Fig. 14d). For the irrigated Mead data (Treatments 1–6), most of the models agreed with one another and with the observations. They also did well for four of the Mead rainfed years, but underestimated Treatments 7 and 12. They did less well with the Bushland data, especially underestimating the SDI irrigation grain yields. The underestimation of SDI grain yields is likely tied to overly large partitioning of applied water to soil E, leaving less available water for T and grain yield formation. Many models, including DSSAT, lack true SDI capability and applied the water to the soil surface in this study. Because SDI was more efficient in water use than the MESA irrigation method in the actual fields used for this study (Evett et al., 2020) and, therefore, likely will be more widely used in the future, the inability to handle SDI is an emerging lacuna in many of the models that should be addressed in future.

3.6.2. Inter-comparisons of grain yield within the DSSAT and other model families

DCFS was the best of the several model flavors within the DSSAT family to simulate grain yield, as indicated by nRMSE for Phase 4 (Fig. 9e). However, general patterns are not obvious in Fig. 9e. Nevertheless, some patterns emerge from a direct comparison in Fig. 10e. IXIM was slightly better than CERES. FAO-56 emerged as the best ETp method followed by Priestly-Taylor and then ASCE standardized reference ET equation with grass (short, 12 cm) crop coefficients and then alfalfa coefficients (tall, 50 cm) (Fig. 10e), which might be somewhat biased because they were not independently calibrated. There was no

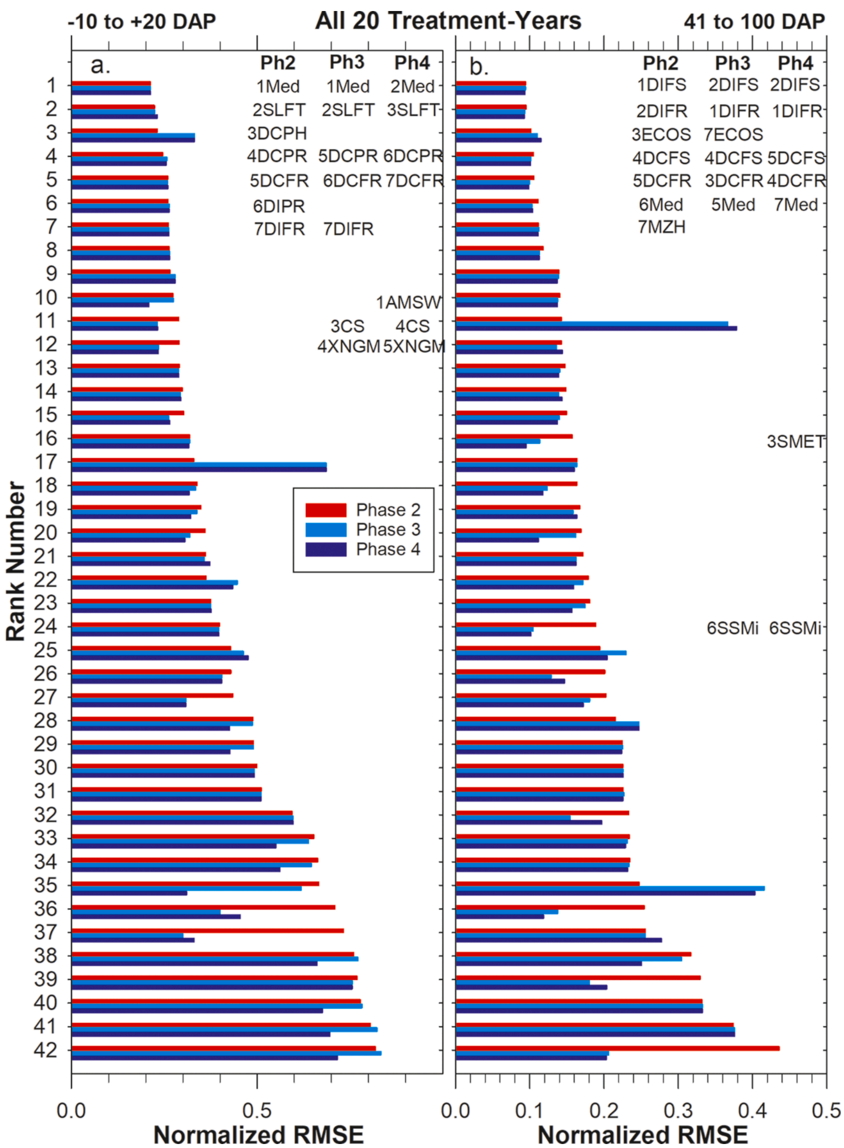

Fig. 13. (a.) Normalized root mean squared error (nRMSE) between observed and simulated cumulative ET values from −10 to +20 days after planting (DAP)(mostly soil E) for all 20 treatment-years for all the models. Phases 2, 3, and 4 are identified by red, cyan, and blue bars with Phase 2 at the top and Phase 4 at the bottom of each group. The models have been sorted in ascending order of nRMSE for Phase 2 from top to bottom of the graph with the rank numbers on the left axis indicating their ranking for Phase 2. The Median (Med) and the six best models (lowest nRMSE) for Phase 2 are listed under "Ph2". Somewhat similarly, the Median and six best models for Phases 3 and 4 are also listed under "Ph3", and "Ph4", but because the modelers made different adjustments going from phase to phase, their rank order changed, so the names along with their nRMSE rank are in different positions down the graph. Phase 1 is not shown because of the planting density error for the six irrigated Maize treatment-years. (b.) Same as for (a.) except the data are for 41 to 100 DAP (mostly crop canopy T).

significant difference in the ability to simulate grain yield between the two methods for simulating soil E.

There was little difference in grain yield simulation ability between the two flavors of the STICS model or of the MAIZSIM model (Fig. 11e). However, XNSM tended to be better than XNGM, although XNGM uses a "more physiological" approach to simulate growth based on the principle of functional balance, in contrast to XNSM, in which a more or less predetermined scheme is used for partitioning of photosynthates.

3.7. K-means clusters

In Figs. 15a and 15c the nRMSE of simulated grain yields for 40 of the models (plus their medians) and of simulated biomass accumulation for 39 of the models (plus their medians), respectively, are compared against the nRMSE of the simulated cumulative ETs for the −10 to +20 DAP time period (which was mostly Es for these mostly bare soil conditions). These graphs show that for many of the models the relative errors for simulating ETs tended to be larger than those for biomass and grain yield, which is consistent with the survey of Seidel et al. (2018) who found that few modelers calibrate the ET aspects of their models. Further, k-means clustering analyses with the number of clusters (k) specified to be four, the models were grouped into the four clusters illustrated in Figs. 15a and 15c. As can be seen, the k-means program identified a cluster of models that did quite well with the nRMSE for grain yields and biomass less than about 0.25 and that for ETs less than 0.35. One of the other clusters did poorly at simulating grain yield and biomass, and the other two clusters did progressively worse at simulating ETs. Figs. 15a and 15c suggest that a model's ability to simulate ETs well early in the growing season from −10 to +20 DAP can carry on through the seasons to help simulate biomass and grain yields well too.

Similarly, Fig. 15b and 15d illustrate the nRMSEs for grain yield and biomass against the nRMSE for the cumulative ETs from 41 to 100 DAP, when there were mostly full crop canopies. Comparing Fig. 15b with 15a and comparing Fig. 15d with 15c, it is apparent that the models were better at simulating the cumulative ETs for the full canopies than they were for bare soil at the beginning of the growing seasons. Again, k-means cluster analyses identified clusters of models that did quite well at simulating grain yields, biomass, and full canopy ETs quite well with the nRMSE of grain yield, biomass, and ETs all less than about 0.2. It is not surprising that there is such a cluster of models that can simulate ETs well during midseason which aids them to also simulate biomass and grain yields well.

Table 2 lists the models included in the best-performing clusters in Fig. 15. There is overlap among the four categories, but CS, AMSW,

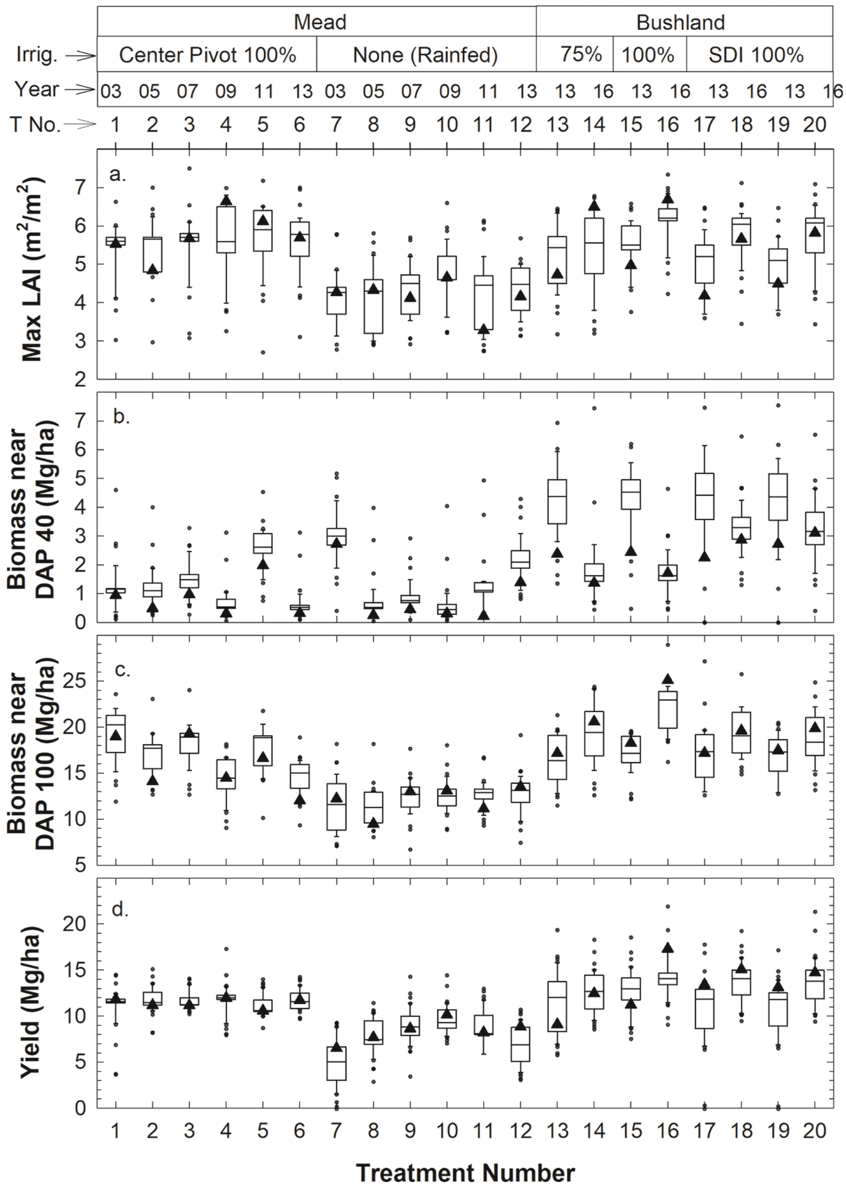

Fig. 14. Box plots for Phase 4 of (a) maximum leaf area index, (b) biomass at about 40 days after planting (DAP), (c) biomass at about 100 DAP, and (d) final grain yield for all 20 treatment-years. Also shown are the corresponding observations (triangles).

ECOS, XNSM, and AHC all excelled enough to appear in all four. Similarly performing well enough to appear in all four categories are three flavors from the DSSAT family: DIFR, DCFR, and DIGR. Not surprisingly, the ensemble median did very well, being first or second in all the categories, consistent with previous inter-comparisons, e.g., Asseng et al. (2015). Among these eight models, CS, XNSM, AHC, DIFR, DCFR, and DIGR all use FAO-56 (Allen et al., 1998) to compute ETp (Table S1). ETo was used as ETp for DIFR and DCFR, whereas DIGR used crop coefficients to adjust ETo to ETp and then various simulated or calculated crop or energy extinction coefficients were used to obtain ETs. AMSW simulates T using the transpiration efficiency approach and E using Ritchie's (Ritchie, 1972) two-stage method (Probert et al., 1998; Keating et al., 2003). ECOS simulates ETs from net radiation that is partitioned into latent, sensible, and soil heat fluxes with energy balances on the canopy and soil surfaces approach (Grant et al., 2007; Grant and Flanagan, 2007). DIGR uses the ASCE Standardized "Short Crop" (12-cm grass) ETo (R.G. Allen et al., 2005), which is a successor to FAO-56, with maize crop coefficients computed from simulated LAI to adjust ETo to ETp. Thus, six of these models have similar core approaches for simulating ETs but differ in other ways such as partitioning to leaf area or soil moisture movement, etc. AMSW and ECOS are both unique in their own ways within this elite group. The three DSSAT models all use the Ritchie-two-stage method for soil water evaporation rather than the Sulieman method, highlighting the need for E methods with improved upward movement of soil water and more accurate E loss in the incomplete canopy phase.

However, something all eight models have in common is that they all have been widely used for a long time under a wide range of conditions. This includes the lesser-known XNSM because it is a hybrid model with elements from both the CERES model (Jones and Kiniry, 1986) and the SUCROS model family (van Laar, 1992; Wang and Engel, 2000). AHC is also included because it was developed based on a coupling of the significantly modified SWAP model (van Dam et al., 1997) and the EPIC crop growth model (Williams et al., 1989). Thus, there has been time for several generations of modelers to improve these models so that they perform well over a wide range of climatic and soil conditions.

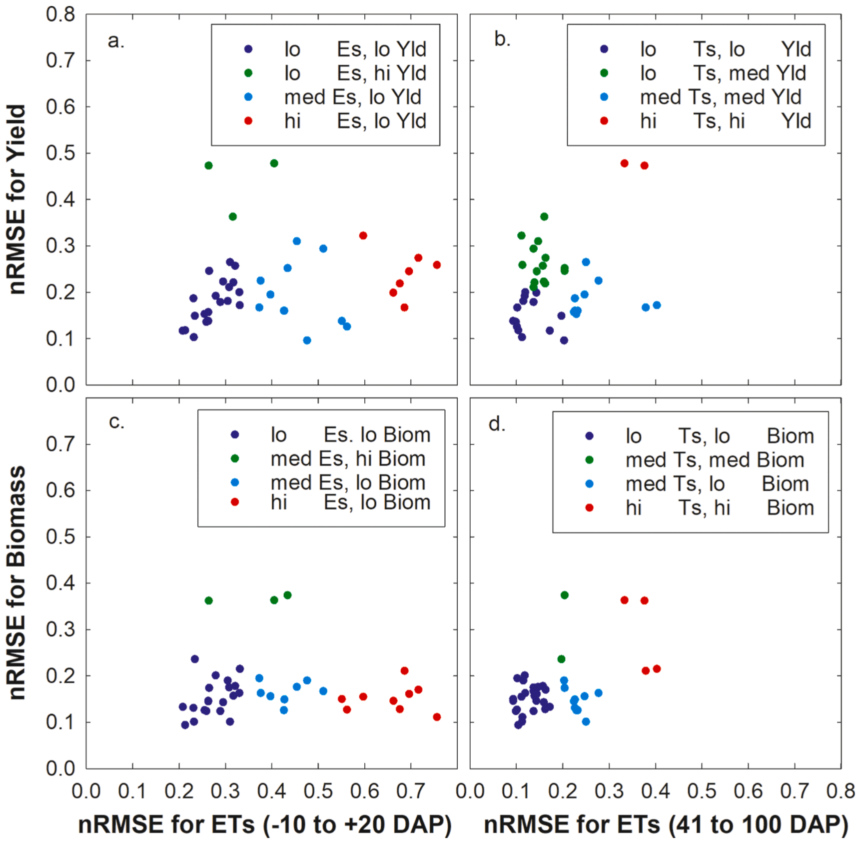

Fig. 15. (a.) K-means clusters of the nRMSE for grain yields of 41 models (plus their median) for all 20 treatments versus the corresponding ETs for −10 to +20 DAP (mostly Ea). (b.) Same as (a) but for the 41 to 100 DAP periods (mostly Ta). (c.) K-means clusters of the nRMSE for biomass accumulation of 40 models (plus their median) from 41 to 100 DAP versus the corresponding ETs for from −10 to +20 DAP (mostly Ea). (d.) Same as (c) but for the 41 to 100 DAP periods (mostly Ts). (Note: one of the 41 models did not simulate grain yield and two did not simulate biomass.).

Table 2
Lists of models in Fig. 15 identified as being in the K-means clusters of best models (lowest nRMSE) for Phase 4 for simulated grain yields and biomass versus lowest nRMSE for simulated ETs for −10 to +20 DAP (mostly bare soil, Es) and 41 to 100 DAP (mostly closed canopy, Ts). The models are ranked according to their sums of nRMSE for grain yield or biomass plus that for ETs.

Ranking	Yield vs. Es	Yield vs. Ts	Biomass vs Es	Biomass vs Ts
1	AMSW	Med	CS	Med
2	Med	CS	Med	CS
3	CS	AMSW	DCFS	DCFR
4	XNGM	SLFT	DIFR	SWB
5	DCFR	DCPR	DIFS	DCFS
6	DIFR	DCFR	DCFR	DIFR
7	DCPR	DIPR	SSMi	DIFS
8	SLFT	DIFR	AMSW	DIGR
9	DIPR	SLUS	ECOS	MZH
10	DIGR	DIGR	DACT	AHC
11	XNSM	DCGR	XNSM	DIGS
12	ECOS	BIOM	DIGR	DCGS
13	DCPH	XNGM	AHC	DIAR
14	BIOM	DIAR	DIGS	SSMi
15	DCGR	XNSM	XNGM	DCGR
16	AQCP	AQCP		MZD
17	AHC	AHC		AMSW
18	DIAR	ECOS		ECOS
19	SLUS	DCAR		DIAS
20	DCAR	DCRH		AQCP
21				XNSM
22				TMOD
23				DCAS
24				DCAR

4. Conclusions with discussion

4.1 Like the previous maize model ET inter-comparison (Kimball et al., 2019), again there was wide variability among the models in their ability to simulate ET, both on daily and on longer interval bases. The variability generally persisted even as the modelers received more information going from one phase to another, although a few modelers did make performance improvements.

4.2 Being among the best models at simulating daily ETs did not guarantee that a model would be among the best at simulating cumulative ETs.

4.3 Nevertheless, eight models, as well as the ensemble median, were identified that did well at simulating (a) cumulative ETs from −10 to +20 DAP (mostly soil E), (b) cumulative ETs from 40 to 100 DAP (mostly canopy T), (c) biomass accumulation, and (d) final grain yield. The models were CS, AMSW, ECOS, XNSM, AHC, DIFR, DCFR, and DIGR. Six of them follow the general approach of using FAO-56/Penman-Monteith (Allen et al., 1998; R.G. 2005) to simulate ETs, while AMSW uses a transpiration efficiency approach (Probert et al., 1998; Keating et al., 2003), and ECOS uses an energy balance approach (Grant et al., 2007; Grant and Flanagan, 2007). All of these models or their ancestors have been in existence and have been widely used for a long time. Thus, there has been time for improvement over a wide range of climatic and soil conditions. Unlike the previous inter-comparison (Kimball et al., 2019), none of the simpler models were among the best at simulating all four variables for this study involving a wider range of environmental conditions from two locations.

4.4 Although the ensemble median was not among the best estimates of soil moisture (Supplementary), it was at the top or close to the top for all other categories. That the ensemble median generally outperforms any individual model is consistent with previous intercomparisons, e.g., Asseng et al. (2015).

4.5 Within the DSSAT family, the older Ritchie (1972) approach for simulating soil E was markedly better than the newer Suleiman and Ritchie (2003, 2004) approach, which appeared to overestimate upward movement of soil moisture.

4.6 Further, within the DSSAT family, the FAO-56 (Allen et al., 1998) method for calculating potential ETp was best for simulating ETs from 40 to 100 DAP (mostly canopy T) and worse for −10 to +20 DAP (mostly soil E). The Priestly and Taylor (1972) method was best for soil E and worse for canopy T. The ASCE Standardized Equation approach with short or tall crop coefficients (R.G. Allen et al., 2005) was intermediate for canopy T and worst for soil E, although this result might be somewhat biased because they were not independently calibrated.
4.7 DSSAT CSM-IXIM tended to be slightly better than DSSAT CSM-CERES for simulating canopy T, probably because IXIM simulated leaf area progression better.
4.8 Both STCK (which considers one surface to compute ETp) and STSW (which considers both soil and canopy surfaces to compute ETp) were among the best models to simulate ETs at the beginning of the seasons, with slightly better results for STSW. During the mid-season periods, STCK globally performed better than STSW, but both performed poorly with SDI in 2013, which distorted results.
4.9 XNSM and XNGM appeared to do equally well at simulating both soil E and canopy T, with XNGM following the more detailed Farquhar modeling approach in calculating photosynthesis and leafT, but greatly simplifying vertical root distribution. However, XNSM did better than XNGM at simulating grain yield, possibly due to its simpler but more robust approach in simulating assimilate distribution among plant organs.

4.9 MZD and MZH both have hourly time steps, yet MZD which uses daily weather data did slightly better than MZH which uses hourly weather data at simulating soil E, but there was no significant difference between them at simulating canopy T. This is somewhat surprising, but nevertheless shows that simulated diurnal patterns of hourly weather can be as accurate as using the actual hourly observations for input to crop growth models with hourly time steps.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

We appreciate access to the comprehensive dataset from Mead, Nebraska, USA, which was collected by the following scientists: Shashi B. Verma, Achim Dobermann, Kenneth G. Cassman, Daniel T. Walters, Johannes M. Knops, Timothy J. Arkebauer, George G. Burba, Brigid Amos, Haishum Yang, Daniel Ginting, Kenneth G. Hubbard, Anatoly A. Gitelson, and Elizabeth A. Walter-Shea. The dataset was collected with support from the DOE-Office of Science (BER: Grant Nos. DE-FG03–00ER62996 and DE-FG02–03ER63639), DOE-EPSCoR (Grant No. DE-FG02–00ER45827), and the Cooperative State Research, Education, and Extension Service, US Department of Agriculture (Agreement No. 2001–38700–11092). Funding was also provided by the National Multidisciplinary Laboratory for Climate Change, RRF-2.3.1–21–2022–00014 project. Additional support was provided by grant "Advanced research supporting the forestry and wood-processing sectoŕs adaptation to global change and the 4th industrial revolution", No. CZ.02.1.01/0.0/0.0/16_019/0000803 financed by OP RDE. KW was supported by the Met Office Hadley centre Climate Programme funded by BEIS.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.agrformet.2023.109396.

References

Allen, R.G., Pereira, L.S., Raes, D., Smith, M., 1998. Crop Evapotranspiration: Guidelines for Computing Crop Water Requirements. Food and Agriculture Organization of the United Nations, Rome, Italy. FAO Irrigation and Drainage Paper 56.
Allen, R.G., Walter, I.A., Elliott, R., Howell, T., Itenfisu, D., Jensen, M., Snyder, R.L., 2005. The ASCE Standardized Reference Evapotranspiration Equation. American Society of Civil Engineers, Reston, Virginia, p. 195.
Annandale, J.G., Benade, N., Jovanovic, N.Z., Steyn, J.M., Du Sautoy, N., 1999. Facilitating irrigation by means of the soil water balance model. Water Research Commission, WRC Report No. 753/1/99, ISBN No. 1 86845 559 9, Pretoria, South Africa.
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, L., 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., I Shcherbak, I., Steduto, P., Stöckle, C., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Travasso, M., Waha, M.K., Wallach, D., White, J.W., Williams, J.R., Wolf, J., 2013. Uncertainties in simulating wheat yields under climate change. Nature Clim. Change 3, 827–832.
Asseng, S., Ewert, F., Martre, P., Rotter, R.P., Lobell, D.B., Cammarano, D., Kimball, B.A., Ottman, M.J., Wall, G.W., White, J.W., Reynolds, M.P., Alderman, P.D., Prasad, P.V. V., Aggarwal, P.K., Anothai, J., Basso, B., Biernath, C., 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., Muller, C., Naresh Kumar, S., Nendel, C., O/'Leary, G., Olesen, J.E., Palosuo, T., Priesack, E., Eyshi Rezaei, E., Ruane, A.C., Semenov, M.A., Shcherbak, I., Stockle, C., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Thorburn, P.J., Waha, K., Wang, E., Wallach, D., Wolf, J., Zhao, Z., Zhu, Y., 2015. Rising temperatures reduce global wheat production. Nature Clim. Change 5, 143–147.
Basso, B., Ritchie, J.T., 2015. Simulating crop growth and biogeochemical fluxes in response to land management using the SALUS model. In: Hamilton, S.K., Doll, J.E., Robertson, G.P. (Eds.), The Ecology of Agricultural Landscapes: Long-Term Research On the Path to Sustainability. Oxford University Press, New York, New York, USA, pp. 252–274.
Bassu, S., Brisson, N., Durand, J.-L., Boote, K., Lizaso, J., Jones, J.W., Rosenzweig, C., Ruane, A.C., Adam, M., Baron, C., Basso, B., Biernath, C., Boogaard, H., Conijn, S., Corbeels, M., Deryng, D., DeSanctis, G., Gayler, S., Grassini, P., Hatfield, J., Hoek, S., Izaurralde, C., Jongschaap, R., Kemanian, A.R., Kersebaum, K.C., Kim, S.-H., Kumar, N.S., Makowski, D., Mueller, 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? Global Change Biol. 20, 2301–2320. https://doi.org/10.1111/gcb.12520.
Best, M.J., Pryor, M., Clark, D.B., Rooney, G.G., Essery, Ménard, C.B., Edwards, J.M., Hendry, M.A., Porson, A., Gedney, A.N., Mercado, L.M., Sitch, S., Blyth, E., Boucher, O., Cox, P.M., Grimmond, C.S.B., 2011. The Joint UK Land Environment Simulator (JULES), model description Part 1: energy and water fluxes. Geosci. Model Dev. 4 (3), 677–699. https://doi.org/10.5194/gmd-4-677-2011 (01 September 2011).
Brisson, N., Perrier, A., 1991. A semi-empirical model of bare soil evaporation for crop simulation models. Water Resour. Res. 27, 719–727.
Brisson, N., Itier, B., L'Hotel, J.C., Lorendeau, J.Y., 1998. Parameterisation of the Shuttleworth-Wallace model to estimate daily maximum transpiration for use in crop models. Ecol. Model. 107, 159–169.
Brisson, N., Gary, C., Justes, E., Roche, R., Mary, B., Ripoche, D., Zimmer, D., Sierra, J., Bertuzzi, P., Burger, P., Bussière, F., Cabidoche, Y.M., Cellier, P., Debaeke, P., Gaudillère, J.P., Hènault, C., Maraux, F., Sequin, B., Sinoquet, H., 2003. An overview of crop model STICS. Eur. J. of Agron. 18, 309–332.
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 CO_2. Field Crops Res 198, 80–92. https://doi.org/10.1016/j.fcr.2016.08.015.
Constantin, J., Willaume, M., Murgue, C., Lacroix, B., Therond, O., 2015. The soil-crop models STICS and AqYield predict yield and soil water content for irrigated crops equally well with limited data. Agric. For. Meteorol. 206, 55–68.
DeJonge, K.C., Thorp, K.R., 2017. Standardized reference evapotranspiration and dual crop coefficient approach in the DSSAT cropping system model. Trans. ASABE 60 (6), 1965–1981.
Doorenbos, J., Pruitt. W.O., 1985. Guidelines for predicting crop water requirements. FAO Irrig. and Drain. Paper 24. FAO, Rome.

Durand, J.L., Delusca, K., Boote, K.J., Lizaso, J., Manderscheid, R., Weigel, H.J., Ruane, A.C., Rosenzweig, C., Ahuja, L., Anapalli, S., Basso, B., Baron, C., Bertuzzi, P., Deryng, D., Ewert, F., Gaiser, T., Gayler, S., Heinlein, F., Kersebaum, F.C., Kim, S.H., Muller, C., Nendel, C., Olioso, A., Priesack, E., Villegas, J.R., Ripoche, D., Seidel, S.I., Srivastava, A., Tao, F., Timlin, D., Twine, T., Wang, E., Webber, H., Zhao, Z., 2018. How accurately do maize crop models simulate the interactions of atmospheric CO_2 concentration levels with limited water supply on water use and yield? Eur. J. Agron. (in press).

Evett, S.R., Heng, L.K., Moutonnet, P., Nguyen, M.L. (Eds.), 2008. Field Estimation of Soil Water content: A practical Guide to methods, instrumentation, and Sensor Technology. International Atomic Energy Agency, Vienna, Austria.

Evett, S.R., Kustas, W.P., Gowda, P.H., Anderson, M.C., Prueger, J.H., Howell, T.A., 2012. Overview of the Bushland evapotranspiration and agricultural remote sensing experiment 2008 (BEAREX08): a field experiment evaluating methods for quantifying ET at multiple scales. Adv. Water. Resour. 50, 4–19. https://doi.org/10.1016/j.advwatres.2012.03.010.

Evett, S.R., Marek, G.W., Colaizzi, P.D., Ruthardt, B.B., Copeland, K.S., 2018a. A subsurface drip irrigation system for weighing lysimetry. Appl. Eng. Agric. 34 (1), 213–221. https://doi.org/10.13031/aea.12597.

Evett, S.R., Marek, G.W., Copeland, K.S., Colaizzi, P.D., 2018b. Quality management for research weather data: USDA-ARS, Bushland, TX. Agrosyst. Geosci. Environ. 1 (1) https://doi.org/10.2134/age2018.09.0036.

Evett, S.R., Marek, G.W., Colaizzi, P.D., Brauer, D.K., O'Shaughnessy, S.A., 2019. Corn and sorghum ET, E, yield, and CWP as affected by irrigation application method: SDI versus mid-elevation spray irrigation. Trans. ASABE 62 (5), 1377–1393. https://doi.org/10.13031/trans.13314.

Evett, S.R., Marek, G.W., Colaizzi, P.D., Brauer, D.K., Howell, T.A., 2020. Are crop coefficients for SDI different from those for sprinkler irrigation application? Trans. ASAB 63 (5), 1233–1242. https://doi.org/10.13031/trans.13920.

Evett, S.R., Copeland, K.S., Ruthardt, B.B., Marek, G.W., Colaizzi, P.D., Howell Sr., T.A., Brauer, D.K., 2022. The Bushland, Texas Maize for Grain Datasets. Ag Data Commons. https://doi.org/10.15482/USDA.ADC/1526317.

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

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

Goudriaan, J., 1977. Crop Micrometerology: A Simulation Study. Simulation Monographs. PUDOC, Wageningen, the Netherlands.

Grant, R.F., Arkebauer, T.J., Dobermann, A., Hubbard, K.G., Schimelfenig, T.T., Suyker, A.E., Verma, S.B., Walters, D.T., 2007. Net biome productivity of irrigated and rainfed maize – soybean rotations: modelling vs. measurements. Agron. J. 99, 1404–1423.

Grant, R.F., Flanagan, L.B., 2007. Modeling stomatal and nonstomatal effects of water deficits on CO_2 fixation in a semiarid grassland. J. Geophys. Res. 112, G03011. https://doi.org/10.1029/2006JG000302.

Hasegawa, T., Lai, T., Yin, X., Zhu, Y., Boote, K., Baker, J., Bregaglio, S., Buis, S., Confalonieri, R., Fugice, J., Fumoto, T., Gaydon, D., Naresh Kumar, S., Lafarge, T., Marcaida, M., Masutomi, Y., Nakagawa, H., Oriol, P., Ruget, F., Singh, U., Tang, L., Tao, F., Wakatsuki, H., Wallach, D., Wang, Y., Wilson, L.T., Yang, L., Yang, Y., Yoshida, H., Zhang, Z., Zhu, J., 2017. Causes of variation among rice models in yield response to CO_2 examined with Free-Air CO_2 Enrichment and growth chamber experiments. Sci. Rep. 7, 14858. https://doi.org/10.1038/s41598-017-13582-y.

Hidy, D., Barcza, Z., Marjanović, H., Ostrogović Sever, M.Z., Dobor, L., Gelybó, G., Fodor, N., Pintér, K., Churkina, G., Running, S., Thornton, P., Bellocchi, G., Haszpra, L., Horváth, F., Suyker, A., Nagy, Z., 2016. Terrestrial Ecosystem Process Model Biome-BGCMuSo v4.0: summary of improvements and new modeling possibilities. Geosci. Model Dev. 9, 4405–4437. https://doi.org/10.5194/gmd-9-4405-2016.

Hoogenboom, G., Porter, C.H., Boote, K.J., Shelia, V., Wilkens, P.W., Singh, U., White, J. W., Asseng, S., Lizaso, J.I., Moreno, L.P., Pavan, W., Ogoshi, R., Hunt, L.A., Tsuji, G. Y., Jones, J.W., 2019a. The DSSAT crop modeling ecosystem. In: Boote, K. (Ed.), Advances in Crop Modelling for a Sustainable Agriculture. Burleigh Dodds Science Publishing, Cambridge, UK. ISBN: 978 1 78676 240 5. www.bdspublishing.com.

Hoogenboom, G., Porter, C.H., Shelia, V., Boote, K.J., Singh, U., White, J.W., Hunt, L.A., Ogoshi, R., Lizaso, J.I., Koo, J., Asseng, S., Singles, A., Moreno, L.P., Jones, J.W., 2019b. Decision Support System for Agrotechnology Transfer (DSSAT) Version 4.7. DSSAT Foundation, Prosser, Washington. http://dssat.net.

Jones, C.A., Kiniry, J.R., 1986. CERES-Maize: A Simulation Model of Maize Growth and Development. Texas A&M University Press, College Station, Texas, p. 194.

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 system simulation. Europ. J. Agron. 18, 267–288.

Kimball, B., Boote, K., Hatfield, J.L., Ahuja, L.R., Stockle, C., Archontoulis, S., Caron, C., Basso, B., Bertuzzi, P., Constantin, J., Deryng, D., Dumont, B., Durand, J., Ewert, F., Gaiser, T., Gayler, S., Hoffmann, M.P., Jiang, Q., Kim, S., Lizaso, J., Moulin, S., Nednel, C., Parker, P., Palosuo, T., Priesack, E., Qi, Z., Srivastava, A., Tommaso, S., Tau, F., Thorp, K.R., Timlin, D.J., 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.

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, F., 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. Global Change Biol 21, 1328–1341.

Liu, B., Asseng, S., Müller, C., Ewert, F., Elliott, J., Lobell, D.P., Matre, P., Ruane, A.C., Wallach, D., Jones, J.W., Rosenzweig, C., Aggarwal, P.K., Alderman, P.D., Anothai, J., Basso, B., Biernath, C., Cammarano, C.D., Challinor, A., Deryng, D., De Sanctis, G., Doltra, J., Fereres, E., Folberth, C., Garcia-Vila, M., Gayler, S., Hoogenboom, G., Hunt, L.A., Izaurralde, R.C., Jabloun, M., Jones, C.D., Kersebaum, K.C., Kimball, B.A., Koehler, A.-K., Kumar, A.-K., Nendel, S.N., O'Leary, C., Olesen, G.J., Ottman, J.E., Palosuo, M.J., Prasad, T., Priesack, P.V.V., Pugh, E., Reynolds, T.A.M., Rezaei, M., Rötter, E.E., Schmid, R.P., Semenov, E., Shcherbak, M.A., Stehfest, I.E., Stöckle, C.O., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Thorburn, P., Waha, K., Wall, G.W., Wang, E., White, J.W., Wolf, J., Zhao, Z., Zhu, Y., 2016. Similar estimates of temperature impacts on global wheat yield by three independent methods. Nat. Clim. Chang 6, 1130–1138, 10.10.38/NCLIMATE3115.

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 (3–4), 235–265. https://doi.org/10.1016/S1161-0301(02)00107-7.

Lopez-Cedron, F.X., Boote, K.J., Pineiro, J., Sau, F., 2008. Improving the CERES-Maize model ability to simulate water deficit effects on maize production and yield components. Agron. J. 100, 296–307.

Maiorano, A., Martre, P., Asseng, S., Ewert, F., Müller, C., Rötter, R.P., Ruane, A.C., Seminov, M.A., Wallach, D., Wang, E., Alderman, P.D., Kassie, B.T., Biernath, C., Basso, B., Cammarano, D., Challinor, A.J., Doltra, J., Dumont, B., Rezaei, E.E., Gayler, S., Kersebaum, K.C., Kimball, B.A., Koehler, A.-K., Liu, B., O'Leary, G.J., Olesen, J.E., Ottman, M.J., Priesach, E., Reynolds, M., Stratonovitch, P., Streck, T., Thorburn, P.J., Waha, K., Wall, G.W., White, J.W., Zhao, Z., Zhu, Y., 2017. Crop model improvement reduces the uncertainty to temperature of multi-model ensembles. Field Crops Res., 202, 5–20.

Mancosu, N., Spano, D., Orang, M., Sarreshteh, S., Snyder, R.L., 2016. SIMETAW#—a model for agricultural water demand planning. Water Resour. Manag. 30, 541–557. https://doi.org/10.1007/s11269-015-1176-7.

Marek, T.H., Schneider, A.D., Howell, T.A., Ebeling, L.L., 1988. Design and construction of large weighing monolithic lysimeters. Trans. ASAE 31 (2), 477–484. https://doi.org/10.13031/2013.30734.

Marek, G.W., Evett, S.R., Gowda, P.H., Howell, T.A., Copeland, K.S., Baumhardt, R.L., 2014. Post-processing techniques for reducing errors in weighing lysimeter evapotranspiration (ET) datasets. Trans. ASABE 57 (2), 499–515. https://doi.org/10.13031/trans.57.10433.

Monteith, J.L., 1965. Evaporation and environment. 19th Symposia of the Society for Experimental Biology, 19. University Press, Cambridge, pp. 205–234.

Moore, C., Berardi, D., Blanc-Betes, E., DeLucia, E.H., Dracup, E.C., Egenriether, S., Gomez-Casanovas, S.N., Hartman, M.D., Hudiburg, T., Kantola, I., Masters, M.D., Parton, W.J., van Allen, R., von Haden, A.C., Wang, W.H., Bernacchi, C.J., 2020. The carbon and nitrogen cycle impacts of reverting perennial bioenergy switchgrass to an annual maize crop rotation. GCB Bioenergy 12, 941–954. https://doi.org/10.1111/gcbb.12743.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830.

Perego, A., Giussani, A., Sanna, M., Fumagalli, M., Carozzi, M., Alfieri, L., Brenna, S., Acutis, M., 2013. The ARMOSA simulation crop model: overall features, calibration and validation results. Ital. J. Agrometeorol. 3, 23–38.

Penman, H.L., 1948. Natural evaporation from open water, bare soil, and grass. Proc. Ro. Soc. London 194, 120–145.

Priesack, E., Gayler, S., Hartmann, H.P., 2006. The impact of crop growth sub-model choice on simulated water and nitrogen balances. Nutrient Cycl. Agroecosyst. 75, 1–13. https://doi.org/10.1007/s10705-006-9006-1.

Priestley, C.H.B., Taylor, R.J., 1972. On the assessment of surface heat flux and evaporation using large-scale parameters. Monthly Weather Rev 100, 81–92.

Probert, M.E.E., Dimes, J.P.P., Keating, B.A.A., Dalal, R.C.C., Strong, W.M.M., 1998. APSIM's water and nitrogen modules and simulation of the dynamics of water and nitrogen in fallow systems. Agric. Syst. 56, 1–28. https://doi.org/10.1016/S0308-521X(97)00028-0.

Ritchie, J.T., 1972. Model for predicting evaporation from a row crop with incomplete cover. Water Resour. Res. 8, 1204–1213.

Sau, F., Boote, K.J., Bostick, W.M., Jones, J.W., Minguez, M.I., 2004. Testing and improving evapotranspiration and soil water balance of the DSSAT crop models. Agron. J. 96, 1243–1257.

Seidel, S.J., Palosuo, T., Thorburn, P., Wallach, D., 2018. Towards improved calibration of models – Where are we now and where should we go? Eur. J. Agron. 94, 25–35. https://doi.org/10.1016/j.eja.2018.01.006.

Shelia, V., Šimůnek, J., Boote, K.J., Hoogenboom, G., 2018. Coupling DSSAT and HYDRUS-1D for simulations of soil water dynamics in the soil-plant-atmosphere system. J. Hydrol. Hydromech. 66 (2), 232–245.

Shuttleworth, W.J., Wallace, J.S., 1985. Evaporation from sparse crops - an energy combination theory. Quart. J. Roy. Meteorol. Soc. 111, 839–855.

Šimůnek, J., Huang, K., van Genuchten, M., 1998. The HYDRUS Code for Simulating the One- Dimensional Movement of Water, Heat, and Multiple Solutes in Variably-

Saturated Media, Version 6.0. U.S. Salinity Lab., United States Dep. of Agriculture, Agricultural Research Service. Tech. Rep. 144.

Šimůnek, J., van Genuchten, M.T., Šejna, M., 2008. Development and applications of the HYDRUS and STANMOD software packages and related codes. Vadose Zone J. 7 (2), 587. https://doi.org/10.2136/vzj2007.0077.

Soltani, A., Sinclair, T.R., 2012. Modeling Physiology of Crop Development, Growth and Yield, 2012. CABI International, p. 322.

Stöckle, C.O., Donatelli, M., Nelson, R., 2003. CropSyst, a cropping systems simulation model. Eur. J. Agron. 18, 289–307.

Suleiman, A.A., Ritchie, J.T., 2003. Modeling soil water redistribution during second-stage evaporation. Soil Sci. Soc. Amer. J. 67, 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, 745–757. https://doi.org/10.1097/01.ss.0000148740.90616.fd.

Suyker, A.E., Verma, S.B., Burba, G.G., Arkebauer, T.J., Walters, D.T., Hubbard, K.G., 2004. Growing season carbon dioxide exchange in irrigated and rainfed maize. Agric. For. Meteorol. 124, 1–13.

Suyker, A.E., Verma, S.B., Burba, G.G., Arkebauer, T.J., 2005. Gross primary production and ecosystem respiration of irrigated maize and irrigated soybean during a growing season. Agric For Meteorol 131, 180–190.

Suyker, A.E., Verma, S.B., 2008. Interannual water vapor and energy exchange in an irrigated maize-based agroecosystem. Agric. For. Meteorol. 148, 417–427.

Suyker, A.E., Verma, S.B., 2009. Evapotranspiration of irrigated and rainfed maize-soybean cropping systems. Agric. For. Meteorol. 149, 443–452.

Thorp, K.R., Marek, G.W., DeJonge, K.C., Evett, S.R., 2020. Comparison of evapotranspiration methods in the DSSAT Cropping System Model: II. Algorithm performance. Comput. Electron. Agric. 177, 105679.

van Dam, J.C., Huygen, J., Wesseling, J.G., Feddes, R.A., Kabat, P., van Walsum, P.E.V., et al., 1997. Theory of SWAP Version 2.0. Simulation of Water Flow, Solute Transport and Plant Growth in the Soil–Water–Atmosphere–Plant Environment, Department of Water resources, WAU, Report 71, Technical Document 45. DLO Winand Staring Centre-DLO.

van Laar, H.H., Goudriaan, J., and van Keulen, H. (eds), 1992. Simulation of crop growth for potential and water-limited production situations (as applied to spring wheat). Simulation Reports CABO-TT, no. 27, Wageningen, 72 pp.

Villalobos, F.J., Fereres, E., 1990. Evaporation measurements beneath corn, cotton, and sunflower canopies. Agron. J. 82, 1152–1159.

Wang, E., Engel, T., 2000. SPASS: a generic process-oriented crop model with versatile windows interfaces. Environ. Modell. Softw. 15, 179–188.

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 crop yield projections is reduced by improved temperature response functions. Nat. Plants 3 (1702), 1–11. https://doi.org/10.1038/nplants.2017.102.

Williams, J.R., Jones, C.A., Kiniry, J.R., Spanel, D.A., 1989. The EPIC crop growth model. Trans. ASAE 32 (2), 497–511.

Willmott, C.J., 1982. Some comments on the evaluation of model performance. Bull. Am. Meteorol. Soc. 63 (11), 1309–1313.

Wolf, J., 2012. User Guide for LINTUL5: Simple Generic Model for Simulation of Crop Growth Under Potential, Water Limited and Nitrogen, Phosphorus and Potassium Limited Conditions. Wageningen University.

Xu, X., Sun, C., Neng, F.T., Fu, J., Huang, G.H., 2018. AHC: An integrated numerical model for simulating agroecosystem processes-Model description and application. Ecological Modelling 390, 23–39.

Yang, Y., Kim, S.-H., Timlin, D.J., Fleisher, D.H., Quebedeaux, B., Reddy, V.R., 2009. Simulating canopy evapotranspiration and photosynthesis of corn plants under different water status using a coupled MaizeSim+2DSOIL Model. Trans. ASAEB 52 (3), 1011–1024.