

Simulation of maize evapotranspiration: An inter-comparison among 29 maize models



Bruce A. Kimball^{a,*}, Kenneth J. Boote^b, Jerry L. Hatfield^c, Laj R. Ahuja^d, Claudio Stockle^e, Sotirios Archontoulis^f, Christian Baron^{g,h}, Bruno Bassoⁱ, Patrick Bertuzzi^j, Julie Constantin^k, Delphine Deryng^{l,m}, Benjamin Dumontⁿ, Jean-Louis Durand^o, Frank Ewert^{p,q}, Thomas Gaiser^q, Sebastian Gayler^r, Munir P. Hoffmann^{o,s}, Qianjing Jiang^t, Soo-Hyung Kim^u, Jon Lizaso^v, Sophie Moulin^w, Claas Nendel^p, Philip Parker^x, Taru Palosuo^y, Eckart Priesack^z, Zhiming Qi^t, Amit Srivastava^q, Tommaso Stella^p, Fulu Tao^{y,A}, Kelly R. Thorp^a, Dennis Timlin^B, Tracy E. Twine^C, Heidi Webber^{p,q}, Magali Willaume^k, Karina Williams^D

^a U.S. Arid-Land Agricultural Research Center, USDA-ARS, Maricopa, AZ 85138, United States

^b University of Florida, Agronomy Department, Gainesville, FL 32611, United States

^c National Laboratory for Agriculture and the Environment, USDA-ARS, Ames, IA 50010-3120, United States

^d Agricultural Systems Research Unit, USDA, Agricultural Research Service, 2150 Centre Avenue, Bldg. D., Suite 200, Ft. Collins, 80526 CO, United States

^e Biological Systems Engineering, Washington State University, 1935 E. Grimes Way, P.O. Box 646120, Pullman WA 99164-6120, United States

^f Iowa State University, Department of Agronomy, Ames, IA 50010, United States

^g CIRAD, UMR TETIS, F-34398 Montpellier, France

^h TETIS, University Montpellier, AgroParisTech, CIRAD, CNRS, IRSTEA, Montpellier, France

ⁱ Michigan State University, Department Geological Sciences and W.K. Kellogg Biological Station, 288 Farm Ln, 307 Natural Science Bldg., East Lansing, MI 48824, United States

^j US1116 AgroClim, INRA 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

^k AGIR, Université de Toulouse, INRA, INPT, INP- EI PURPAN, 24 Chemin de Borde Rouge - Auzeville CS 52627, CastanetTolosan, France

^l Leibniz Centre for Agricultural Landscape Research (ZALF), MuCheberg 15374, Germany

^m IRI THEsys, Humboldt University of Berlin, Berlin, Germany

ⁿ ULg-GxABT, University of Liege, Gembloux Agro-Bio Tech Faculty, Department Agronomy, Bio-Engineering and Chemistry, Crop Science Unit, Passage des Déportés, 2, 5030 Gembloux, Belgium

^o Unité de Recherche Pluridisciplinaire Prairie et Plantes Fourragères, INRA, 86 600 Lusignan, France

^p Leibniz Centre for Agricultural Landscape Research, Eberswalder Straße 84, 15374 Müncheberg, Germany

^q Institute of Crop Science and Resource Conservation, University of Bonn, Katzenburgweg 5D-53115 Bonn, Germany

^r University of Hohenheim, Institute of Soil Science and Land Evaluation, Biogeophysics, Emil-Wolff-Str. 27, 70599 Stuttgart, Germany

^s Georg-August-Universität ät Goettingen, Tropical Plant Production and Agricultural Systems Modeling (TROPAGS), Grisebachstraße 6, Göttingen, Germany

^t Department of Bioresource Engineering, Macdonald Campus, McGill University, 1-024 Macdonald-Steward Hall, Sanite-Anne-de-Belleuve, QC H9X 3V9, Canada

^u University of Washington, School of Environmental and Forest Sciences, Seattle, WA 98195, United States

^v Technical University of Madrid (UPM), Department Producción Agraria-CEIGRAM, Ciudad Universitaria, 28040 Madrid, Spain

^w UMR1114 EMMAH, INRA, 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

^x Spatial Business Integration, Marienburgstrasse 27, D-64297 Darmstadt, Germany

^y Natural Resources Institute Finland (Luke), Latokartanonkaari 9, FI-00790 Helsinki, Finland

^z Helmholtz Center Munich, Institute of Biochemical Plant Pathology, Ingolstaedter Landstr. 1, 85764 Neuherberg, Germany

^A Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Science, 11A, Datun Road, Chaoyang District, Beijing 100101, China

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

E-mail addresses: bruce.kimball@ars.usda.gov (B.A. Kimball), kjboote@ufl.edu (K.J. Boote), Jerry.Hatfield@ars.usda.gov (J.L. Hatfield), laj.ahuja@ars.usda.gov (L.R. Ahuja), stockle@wsu.edu (C. Stockle), sarchont@iastate.edu (S. Archontoulis), christian.baron@cirad.fr (C. Baron), basso@msu.edu (B. Basso), patrick.bertuzzi@inra.fr (P. Bertuzzi), julie.constantin@toulouse.inra.fr (J. Constantin), delphine.deryng@mail.mcgill.ca (D. Deryng), Benjamin.Dumont@ulg.ac.be (B. Dumont), jean-louis.durand@inra.fr (J.-L. Durand), frank.ewert@uni-bonn.de (F. Ewert), tgaiser@uni-bonn.de (T. Gaiser), sebastian.gayler@uni-hohenheim.de (S. Gayler), munir.hoffmann@zalf.de, mhoffma@gwdg.de (M.P. Hoffmann), qianjing.jiang@mail.mcgill.ca (Q. Jiang), soohkim@uw.edu (S.-H. Kim), jon.lizaso@upm.es (J. Lizaso), sophie.moulin@inra.fr (S. Moulin), nendel@zalf.de (C. Nendel), Parker@spatial-business-integration.com, herrparker@gmail.com (P. Parker), taru.palosuo@luke.fi (T. Palosuo), priesack@helmholtz-muenchen.de (E. Priesack), zhiming.qi@mcgill.ca (Z. Qi), amit.srivastava@uni-bonn.de (A. Srivastava), tommaso.stella@zalf.de (T. Stella), fulu.tao@luke.fi, taofl@igsnr.ac.cn (F. Tao), kelly.thorp@ars.usda.gov (K.R. Thorp), Dennis.Timlin@ars.usda.gov (D. Timlin), twine@umn.edu (T.E. Twine), hwebber@uni-bonn.de (H. Webber), magali.willaume@ensat.fr (M. Willaume), karina.williams@metoffice.gov.uk (K. Williams).

<https://doi.org/10.1016/j.agrformet.2019.02.037>

^B Crop Systems and Global Change Research Unit, USDA-ARS, Beltsville, MD, United States^C University of Minnesota, Department of Soil, Water, and Climate, 1991 Upper Buford Cir, St. Paul, MN 55108, United States^D Hadley Centre, FitzRoy Road, Exeter, Devon EX1 3PB, United Kingdom

ARTICLE INFO

Keywords:

Maize
Simulation
Evapotranspiration
Water use
Model
Yield

ABSTRACT

Crop yield can be affected by crop water use and vice versa, so when trying to simulate one or the other, it can be important that both are simulated well. In a prior inter-comparison among maize growth models, evapotranspiration (ET) predictions varied widely, but no observations of actual ET were available for comparison. Therefore, this follow-up study was initiated under the umbrella of AgMIP (Agricultural Model Inter-Comparison and Improvement Project). Observations of daily ET using the eddy covariance technique from an 8-year-long (2006–2013) experiment conducted at Ames, IA were used as the standard for comparison among models. Simulation results from 29 models are reported herein. In the first “blind” phase for which only weather, soils, phenology, and management information were provided to the modelers, estimates of seasonal ET varied from about 200 to about 700 mm. Subsequent three phases provided (1) leaf area indices for all years, (2) all daily ET and agronomic data for a typical year (2011), and (3) all data for all years, thus allowing the modelers to progressively calibrate their models as more information was provided, but the range among ET estimates still varied by a factor of two or more. Much of the variability among the models was due to differing estimates of potential evapotranspiration, which suggests an avenue for substantial model improvement. Nevertheless, the ensemble median values were generally close to the observations, and the medians were best (had the lowest mean squared deviations from observations, MSD) for several ET categories for inter-comparison, but not all. Further, the medians were best when considering both ET and agronomic parameters together. The best six models with the lowest MSDs were identified for several ET and agronomic categories, and they proved to vary widely in complexity in spite of having similar prediction accuracies. At the same time, other models with apparently similar approaches were not as accurate. The models that are widely used tended to perform better, leading us to speculate that a larger number of users testing these models over a wider range of conditions likely has led to improvement. User experience and skill at calibration and dealing with missing input data likely were also a factor in determining the accuracy of model predictions. In several cases different versions of a model within the same family of models were run, and these within-family inter-comparisons identified particular approaches that were better while other factors were held constant. Thus, improvement is needed in many of the models with regard to their ability to simulate ET over a wide range of conditions, and several aspects for progress have been identified, especially in their simulation of potential ET.

1. Introduction

Crop and agroecosystem models are essential tools widely used to evaluate impacts of climate variability and project impacts of future climate change on agricultural production. To increase confidence in the use of such modeling tools, it is important that the crop models be evaluated and tested for accuracy in predicting crop response to climate factors including rising atmospheric carbon dioxide, warming global temperatures, and variability in rainfall. Groups within the Agricultural Model Inter-comparison and Improvement Project (AgMIP) have been inter-comparing multiple models against each other and against CO₂ and temperature response data [wheat (Asseng et al., 2013, 2015; Cammarano et al., 2016; Maiorano et al., 2017; Wang et al., 2017), maize (Bassu et al., 2014; Durand et al., 2018), rice (Li et al., 2015; Hasegawa et al., 2017), potato (Fleisher et al., 2017)]. Simulated evapotranspiration (ET) processes and crop water balance relative to climate variation are critical aspects that influence the accuracy of crop model predictions, yet relatively few modelers calibrate the ET aspects of their models (Seidel et al., 2018).

A few comparisons among methods to simulate ET have been done previously. Sau et al. (2004) evaluated several ET options with the CROPGRO Fab bean model, by careful comparison to soil water balance, and concluded that the FAO-56 option (Allen et al., 1998) was more accurate than the Priestley-Taylor option (Priestly and Taylor, 1972) or the old FAO-24 option (Doorenbos and Pruitt, 1985). In an inter-comparison among 16 wheat models, Cammarano et al. (2016) found good agreement with observed water use data from four sites around the world and a small range in variability among the models. In contrast, in an inter-comparison among 23 maize models, Bassu et al. (2014) found a very large range of predictions of ET among the models, including variations in the ET response to CO₂ concentration ranging between –10 and +30% at 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, we initiated another maize model inter-comparison study to focus on the ET aspects. For this study, we chose an 8-season-long dataset from Ames, Iowa, USA for which eddy covariance measurements of ET were available, and the results from 29 models (Table 1, Supplementary Table S1) are reported herein.

Approaches for simulating ET in the individual crop models are listed in Supplementary Table S1 and are briefly summarized here. Several daily time-step models predict some type of potential evapotranspiration (ET_p), whereby crop and soil surfaces are presumed to be wet with insignificant resistance to water loss. Some simply use a function of solar radiation and temperature (as in Priestly and Taylor, 1972), whereas others use solar radiation, temperature, windspeed and relative humidity/dewpoint, as well as no longer assuming the surfaces are wet (Penman-Monteith per Monteith, 1965; FAO56 per Allen et al., 1998). Temperature-based approaches for computing ET_p include Hargreaves (1975) and Hamon (1963). Some models use a type of energy extinction coefficient by which the canopy absorbs energy from the solar radiation as an exponential function of leaf area index (LAI) to predict potential canopy transpiration (EP_p = Kep*ET_p), with the remaining energy being absorbed by the soil or residue and contributing to potential soil water evaporation (ES_p). This energy extinction coefficient is typically lower than the extinction coefficient used for absorption of photosynthetically-active radiation for photosynthesis, and may be close to 0.5 as measured experimentally at 0.52 by Villalobos and Fereres (1990) and computed theoretically at 0.46 to 0.53 by Goudriaan (1977) and Goudriaan and Van Laar (1994), and shown to work well in crop models (Sau et al., 2004; Boote et al., 2008). Actual soil water evaporation (ES_a) is often less than potential ES_p, depending on soil water content, soil conductivities to water vapor loss and to liquid movement, residue cover, and windspeed. Actual crop

Table 1

List of models and their acronyms. (For details about the evapotranspiration aspects of each, see Supplementary Table S1, List of Models, their evapotranspiration (ET) characteristics, and pertinent references.)

Model Name	Acronym
Agro-IBIS	AGIB
APSIM-maize, SOILWAT (default) Hoffmann	AMD
APSIM-maize SWIM Archontoulis	AMSA
APSIM-maize SWIM Hoffmann	AMSH
APSIM-maize SOILWAT Archontoulis	AMW
Aqyield	AQY
CropSyst	CS
DSSAT-CERES maize ASCE Ritchie	DCAR
DSSAT-CERES maize ASCE Suleiman	DCAS
DSSAT-CERES maize FAO56 Ritchie	DCFR
DSSAT-CERES maize FAO56 Suleiman	DCFS
DSSAT-CERES maize Priestly-Taylor Ritchie	DCPR
DSSAT-CERES maize Priestly-Taylor Suleiman	DCPS
Expert-N - GECROS - Maize	XNGM
Expert-N - SPASS - Maize	XNSM
IXIM	IXIM
JULES-crop	JUL
MAIZSIM Daily	MZD
MAIZSIM Hourly	MZH
MCWLA	MCW
MONICA	MNCA
PEGASUS	PGSS
RZWQM2	RZWQ
SALUS	SLUS
SARRAH	SARA
SIMPLACE LINTUL5 FAO56 SLIM3 CanopyT	SLFT
SIMPLACE LINTUL5 Hargreaves	SLNH
STICS_ETP_READ_Penman_LUE	STCC
STICS_ETP_SW_calc	STSW

transpiration (E_p) is a function of potential E_p, root-available soil water, and crop conductance to water vapor, as well as other possible factors such as soil salinity or hypoxia under wet conditions. Soil water available for transpiration is simulated quite differently by different crop models. Effects of insufficient soil water to meet transpiration demand cause reduction in dry matter gain and leaf area expansion, and the crop models also vary considerably in how this is done. A few crop and ecosystem models are capable of simulating instantaneous (hourly) energy balance of equilibrium canopy and soil temperature and leaf conductance, and thus do not require computing E_p (Kim et al., 2012; Yang et al., 2009; Yin and van Laar, 2005; Pickering et al., 1995; Twine et al., 2013). Another unique approach is that of the APSIM model, in which potential transpiration is predicted as a function of potential daily dry matter gain, a literature-derived transpiration efficiency, CO₂ concentration, and daily vapor pressure deficit (Tanner and Sinclair, 1983; Soufzadeh et al., 2018). The actual transpiration is reduced to account for root-available soil water. Another approach is that of

Table 2

Dates and days of year (DOY) of planting, anthesis, and maturity of the maize crops for each of the years. Also listed are the number of days from 41 days after planting to 20 days before maturity and the number of days for the whole growing seasons from planting to maturity.

Year	Planting		Anthesis		Maturity		Number of days from 41 after planting to 20 before maturity	Number of days from planting to maturity
	Date	DOY	Date	DOY	Date	DOY		
2006	14 Apr	104	na	na	22 Sep ^a	265 ^a	102	162
2007	11 May	131	na	na	22 Sep ^a	265 ^a	75	135
2008	17 May	138	23 Jul	205	1 Oct	275	78	138
2009	22 Apr	112	15 Jul	196	29 Sep	272	101	161
2010	21 Apr	111	20 Jul	201	7 Sep	250	80	140
2011	4 May	124	5 Jul	186	6 Oct	279	96	156
2012	27 Apr	118	15 Jul	197	14 Sep	258	81	141
2013	19 Apr	139	1 Aug	213	13 Sep	256	58	118

^a Estimated from the average number of days between maturity and harvest for the six years when maturity dates were observed.

irrigation engineers (e.g. Allen et al., 2005) who compute a reference ET (E_T or E_{Tr}) following Penman-Monteith using standard values for canopy conductance characteristic of a short crop (12-cm-tall grass) or a tall crop (50-cm-tall alfalfa) for E_T and E_{Tr}, respectively. Then actual ET is calculated using a crop (or dual crop and soil) coefficient times the reference ET (Allen et al., 2005. Recently DeJonge and Thorp (2017) described an approach to combine the reference ET with dual crop coefficients in a crop growth model. Variations among these several approaches to simulate ET were used by the 29 models in this inter-comparison study.

Therefore, the objectives of this study were to inter-compare the 29 models (Table 1) in their ability to simulate ET, which was determined by actual eddy covariance data, and then to identify those approaches which were most accurate.

2. Materials and methods

2.1. Observed data

The observed data came from a commercial farm (N 42°1', W 93°45', 329 m) located about 10 km southwest of Ames, IA. Two fields with Gleysol soils were involved, which alternated between maize and soybean. One soil was a Clarion fine-loamy, mixed mesic Typic Hapludoll and the other was a Harps fine-loamy, mesic Typic Calciaquoll. The fields were subsurface tile-drained, with tiles at a depth of about 1.5 m. The drains were installed in the early part of the 20th century, and they are positioned across the field to intercept water from the poorly drained soils. Observations of water flow adjacent to the field suggest that these tiles did not have any restrictions to water movement

Each fall the fields were plowed to a depth of 25 cm using a chisel plow with 76.2 cm centers. During each winter, sufficient precipitation was received to increase the soil water content to field capacity and above. Then each spring from 2006 through 2013, as soon as the soil was dry enough to allow heavy equipment to operate in the field, the fields were cultivated to a depth of 10 cm, and a medium-season, non-photoperiod-sensitive maize (*Zea mays* L.) variety was sown in rows spaced 76.2 cm apart at a depth of 3.8 cm with a target planting density of 8.65 plants per m². Anthesis and maturity (black layer) dates were recorded for the last six years of the experiment (Table 2), and these dates were furnished to the modelers at the outset of this exercise. Leaf area index (LAI) was measured weekly with a LI-2200 (Li-Cor Biosciences, Lincoln, Nebraska, USA) at five areas adjacent to the eddy covariance tower. In addition, leaf area was measured on 25 plants at the V2, V6, and VT stages of growth as part of the destructive plant sampling using a Li-Cor LI-3100 leaf area meter. Leaf area was converted to LAI with stand counts obtained with each vegetative sampling. Above-ground biomass was sampled near anthesis during the last six of the years, and final grain yields were determined at the end of

every season. Yields were determined with hand harvests of 1 m lengths of rows in five different areas adjacent to the flux towers and also with combine yield monitors. Yields reported herein are at 0% grain moisture.

In each field, instrumentation to measure energy balance components, including ET, included a permanent eddy-covariance (EC) flux station positioned 1.6 m above the soil surface (or 1.0 m above canopy height with the instruments raised weekly until the canopy height was constant). These units consisted of a fast-response open path H₂O gas analyzer (LI-7500, LICOR Biosciences Inc., Lincoln, NE, USA) and a three-dimensional sonic anemometer (CSAT, Campbell Scientific, Inc., Logan, UT, USA) oriented to the prevailing wind direction (i.e., to the south), four-component net radiometer (CNR1, Kipp and Zonen, Delft, The Netherlands) at 10 m from EC sensors and positioned 1 m above the vegetation surface, and soil heat flux plates (REBS HFT-3, Radiation and Energy Balance Systems, Seattle, WA, USA) at 0.1 m below the soil surface. The separation distance between gas analyzer and sonic anemometer sensors was 0.1 m. Because the EC system is free standing and separated from other sensors or objects, this configuration minimizes potential influence of other sensors (e.g., radiometer) on EC measurements via air-flow distortion. Ancillary measurements in each flux station included soil temperature, soil water content at 10-cm

depth, and surface canopy temperature. Pairs of soil thermocouples (Cu–Co Type T) were placed 2 and 8 cm below the surface adjacent to each soil heat flux plate. Soil water content in the top 0.1 m at each site was measured with Hydra Probe (Stevens Water Monitoring Systems, Portland, OR, USA). Surface canopy temperature was monitored by high-precision infrared radiometric temperature sensors (IRT, 15° field of view, Apogee Instruments, Logan, UT, USA) placed at 0.2 m above the EC sensors and oriented at a 45° from nadir. Signals from all instrumentation were recorded at rates of 10 Hz, and 15-min averages were stored in dataloggers (Campbell Scientific CR5000).

Instruments were calibrated at the beginning and end of each growing season and data quality was checked daily. Energy balance closure was checked for each 15 min interval of data and averaged between 0.85 and 0.90 over a year. Careful data processing was carried out in three sequential steps including corrections, quality control, and gap filling. Turbulent fluxes were corrected following the Webb–Pearman–Leuning algorithm (Webb et al., 1980). Soil heat flux plate data were corrected for heat storage in the top 10-cm soil layer assuming a bulk density of 1.2 Mg m⁻³ and using measured soil temperature and water content (Sauer, 2002). When the soil was frozen, a fractional volumetric ice content of 0.4 and heat capacity of 1.76 MJ m⁻³ °C⁻¹ were assumed for the G storage correction (Hernandez-

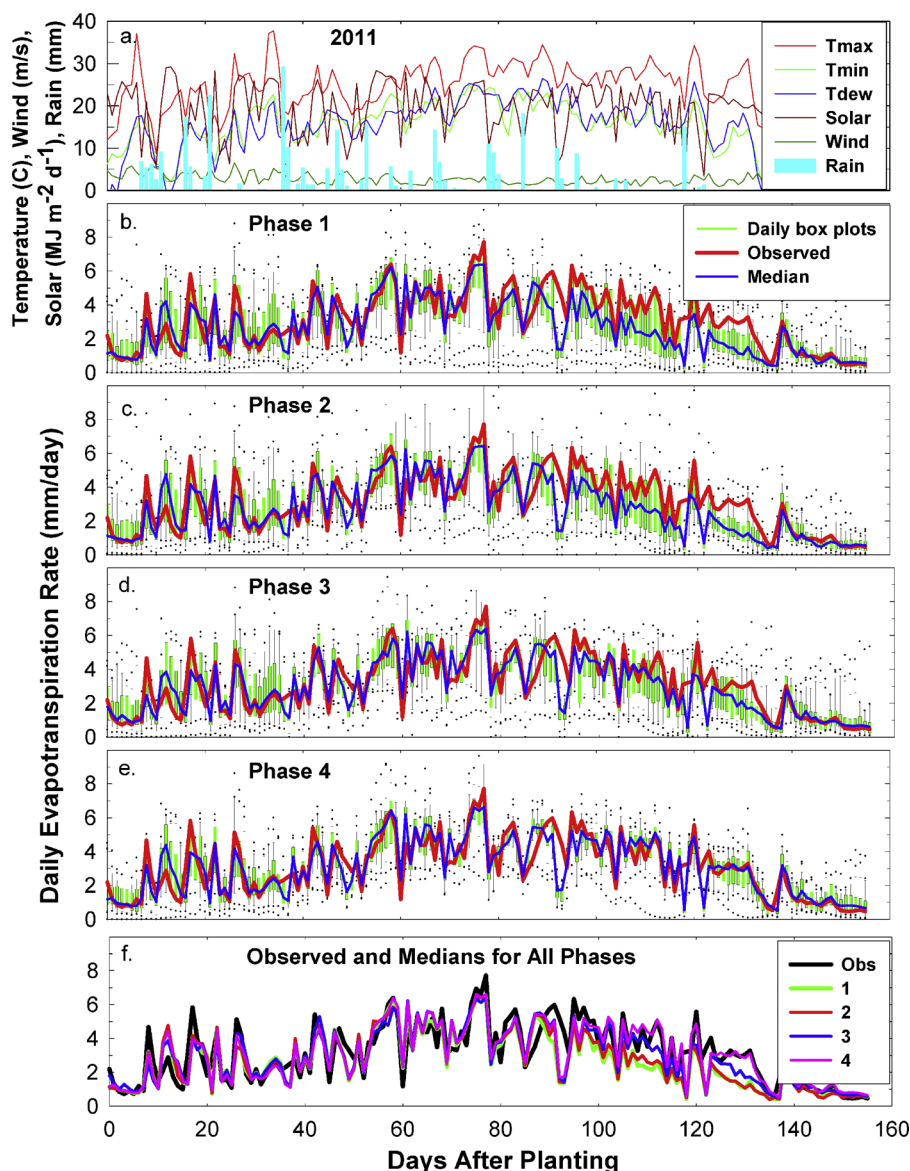


Fig. 1. (a.) Weather variables (maximum and minimum air temperature, dew point, solar radiation, wind speed, rainfall) during 2011, a “typical” rainfall year. (b.) Box plots of daily evapotranspiration (ET) where the lower and upper limits of the box indicate the 25th and 75th percentile of ET values simulated by 29 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 29 models are also shown. The simulated values in this plot came from Phase 1, a “blind” test whereby the modellers were only given weather, phenology, management, and soils information, but no crop response data. (c.) Same as (b.) except for Phase 2 whereby the modellers were given leaf area index data for all eight years. (d.) Same as (c.) except for Phase 3 whereby the modellers were given the observed ET, yield, soil water content at 10 cm, and other data for 2011. (e.) Same as (d.) except for Phase 4 whereby the modellers were given the all the ET, yield, growth, and soil water data for all eight years, as well as options for handling a water table. (f.) Observed daily ET values as well as the median simulated ET values for Phases 1, 2, 3, and 4.

Ramirez et al., 2010; Dold et al., 2016). Quality control and gap filling processes are described in Hernandez-Ramirez et al. (2010) and Dold et al. (2016).

2.2. Simulations

The simulations were conducted by 24 modeling groups from around the world with 29 models completing the inter-comparison. The models are listed in Table 1, and pertinent information about how they compute ET is given in Table S. The study was conducted in four phases:

- 1 “Blind phase.” The modelers were sent key input data about soils, weather, and management (planting dates, fertilizer applications, etc.) information. They also received anthesis and maturity dates for the last six years of the experiment, but no other information about plant growth, yield, or water use.
- 2 “LAI phase”. The modelers were sent time-series leaf area index (LAI) observations for all eight years (about fifteen observations per year). The objective of this phase was to see if improving the simulation of leaf area growth also improved the ET predictions.

- 3 “2011 phase”. The modelers were sent all biomass, yield, daily 10-cm soil moisture, and daily ET data for 2011, which was the most “typical” year for rainfall.
- 4 “All phase”. In this final phase, the modelers were provided with all biomass, yield, daily 10-cm soil moisture, and daily ET data for all the years to allow for detailed calibration of the models.

Lacking any initial soil moisture data at planting time, the modelers were told to assume the soils were at field capacity on day-of-year 91 and to start their simulations then. This start date allowed for several days of equilibration before the actual planting days. However, one result from the first three phases was that in 2012, an infamous drought year (Rippey, 2015), the observed yields and ET were the highest of any other year (suggesting more sun and ample water), whereas all the models simulated very low yields and ET. This discrepancy plus observations of a shallow water table from another close field in another season (Logsdon et al., 2009; Ordóñez et al., 2018) plus the fact the tile drains were at 150 cm and that roots can go below this depth caused us to decide that for Phase 4, the modelers could assume saturation below 1.5 m and that they could take steps to maintain a water table or otherwise provide more water at the deeper depths. However, because

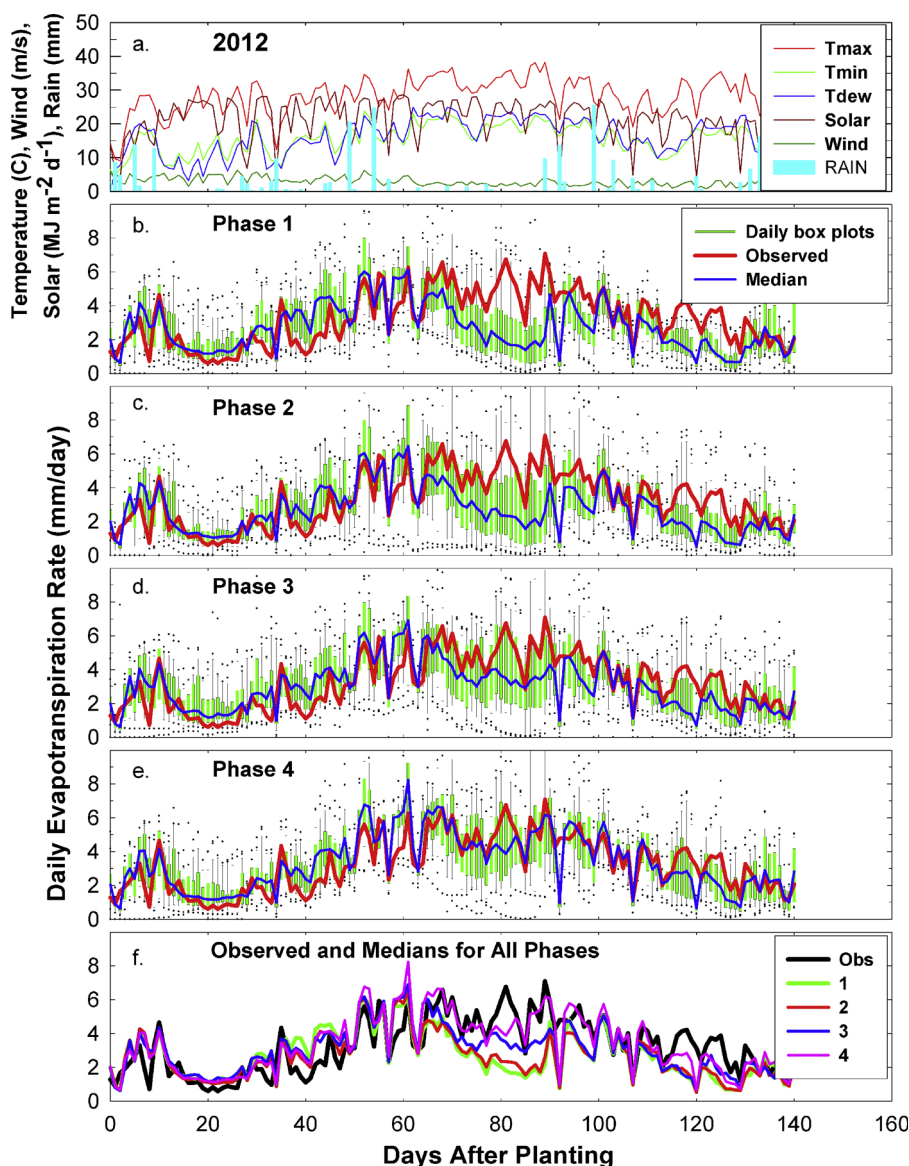


Fig. 2. Same as for Fig. 1 except for 2012, the driest year.

the initial saturation below 150 cm and the water table were both introduced for Phase 4, we do not know which was more effective. Nevertheless, as we will present in the results, this unplanned occurrence indicates that a water table capillary fringe and root access to soil moisture at depth are critical during rainfall-limited seasons.

2.3. Methods for evaluating model performance

Correlation coefficients (r), root mean squared differences between observed and simulated values (RMSQ), average differences, and slopes and intercepts of regressions of observed on simulated data were computed, and along with graphs, they are presented as Supplementary statistical data. However, the mean squared deviation (MSD) method presented by Gauch et al. (2003) was primarily used to evaluate the deviations of simulated estimates from observed values. There are three ways the simulated values can deviate from the observations. First, there can be a standard bias ($SB = (X - Y)^2$, where X is the average of the modeled values and Y is the average of the observations. Second, there can be a non-unity of the regression slope ($NU = (1-b)^2 * (\sum x_n^2/N)$, where b is the slope, x_n is the deviation of the n^{th} modeled value from its mean, and N is the number of observations. Third, there can be a lack of correlation ($LC = 1-r^2) * (\sum y_n^2/N)$, where r is the

correlation coefficient, and y_n is the deviation of the n^{th} observation from its mean. Following Gauch et al., $MSD = \Sigma(X_n - Y_n)^2/N = SB + NU + LC$

3. Simulation results

3.1. Daily ET for a typical year, 2011

The predictions of daily ET for the initial blind Phase 1 for a typical rainfall year (2011) exhibited a six-fold variation among the models (Fig. 1b). However, the box plot from the lowest 25% to the highest 25% of the models was much smaller, although on many days its range was a factor of two. Nevertheless, the median of all the models was fairly close to the observations until about 80 days-after-planting (DAP), when the models tended to be below observations for most of the rest of the growing season. Supplying LAI for all years (Phase 2), gave only slight improvement (Fig. 1c). However, not surprisingly, when the modelers were supplied all the data for 2011 (Phase 3), including daily ET values, there was a closer agreement of the medians to the observations throughout the year, including the latter part of the season after DAP 80 (Fig. 1d). Additionally supplying all the available data for all eight years (Phase 4) brought further improvement (Fig. 1e).

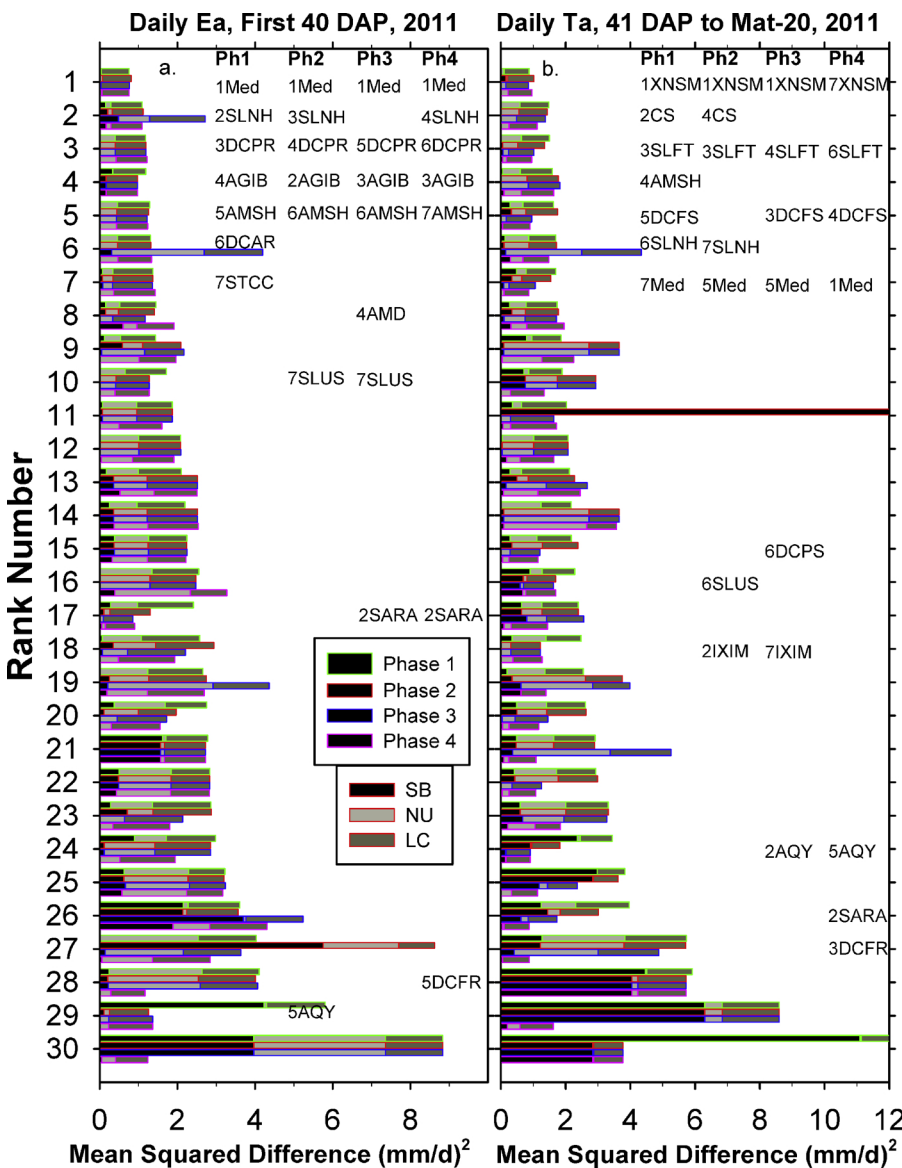


Fig. 3. (a.) Mean squared differences (MSD) between observed and simulated actual daily E_a values for the first 40 days after planting (soil E dominated) for “typical” 2011 growing season for each of maize simulation models. Phases 1, 2, 3, and 4 are identified by green-, red-, blue-, and pink-edge 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 MSDs for Phase 1 (blind) from top to bottom of the graph with the rank numbers on the left axis indicating their ranking for Phase 1. The Median and the six best models (lowest MSD) for Phase 1 are listed under “Ph1”. Somewhat similarly, the Median and six best models for Phases 2 through 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 MSD rank are in different positions down the graph. Further, the bars show the components of MSD: SB for standard bias in medium grey at the right of each bar, NU for non-unity slope in light grey in the middle, and LC for lack of correlation in black at the left. (b.) Same as for (a.) except the data are for the mid-season (crop T dominated) from 41 days after planting to 20 days before observed maturity.

The steadily increasing agreement of the medians to the observations with the progression from Phases 1 to 4 is especially visible at about DAP 110 in Fig. 1f.

3.2. Daily ET for a dry year, 2012

The year 2012 was infamous for causing crop failure in many parts of the Mid-West and further south through the Mississippi River valley (Rippey, 2015). The rainfall record for this Ames, IA dataset shows little rain from about DAP 52 through 90 (Fig. 2a). Surprisingly, observed ET continued at a high level, about 5 mm/day, through this drought period (Fig. 2b–f), while the simulated values dropped steadily for blind Phase 1 (Fig. 2b). Supplying LAI (Phase 2) did not improve agreement with observations (Fig. 2), but supplying all 2011 data (Phase 3) helped somewhat. As discussed in Methods, for Phase 4, modelers were told to allow soil water to be saturated below tile drain level at 150 cm, and they could implement a water table if supported by their model. The upward movement of water from water tables below the root zone recharged by winter and spring rains has been observed for similar soils in this area during droughts (Logsdon et al., 2009) and elsewhere in the Corn Belt (Rizzo et al., 2018). Consequently, Phase 4 simulated median values were able to approach observations (Fig. 2e, f). Different from

2011, most models over-estimated ET during the first 40 days (dominated by soil water evaporation) during all four phases. This difference between typical versus dry years would indicate that soil evaporation during dry periods is generally over-estimated by the models.

3.3. Performance of individual models to predict daily ET for typical 2011 and dry 2012

Because soil surface evaporation, E, and plant transpiration, T, are such different processes, yet both are measured together in ET, it was decided to split the growing season into two parts to test model ET performance. Therefore the first 40 days after planting when the seeds are germinating and the canopy is emerging and developing (time of dominant E) was separated from the period 41 days after planting until 20 days before observed maturity (time of dominant T). For the first 40 DAP in 2011, there was little variation among the better (lowest MSD) models for Phase 1 (Fig. 3a), and there was little improvement or change for successive phases for many of the models. In contrast, for the mid-season T-dominated period, there was more variation among the models and many showed improvement with successive Phases (Fig. 3b).

Some individual modelers' adjustments or lack thereof are

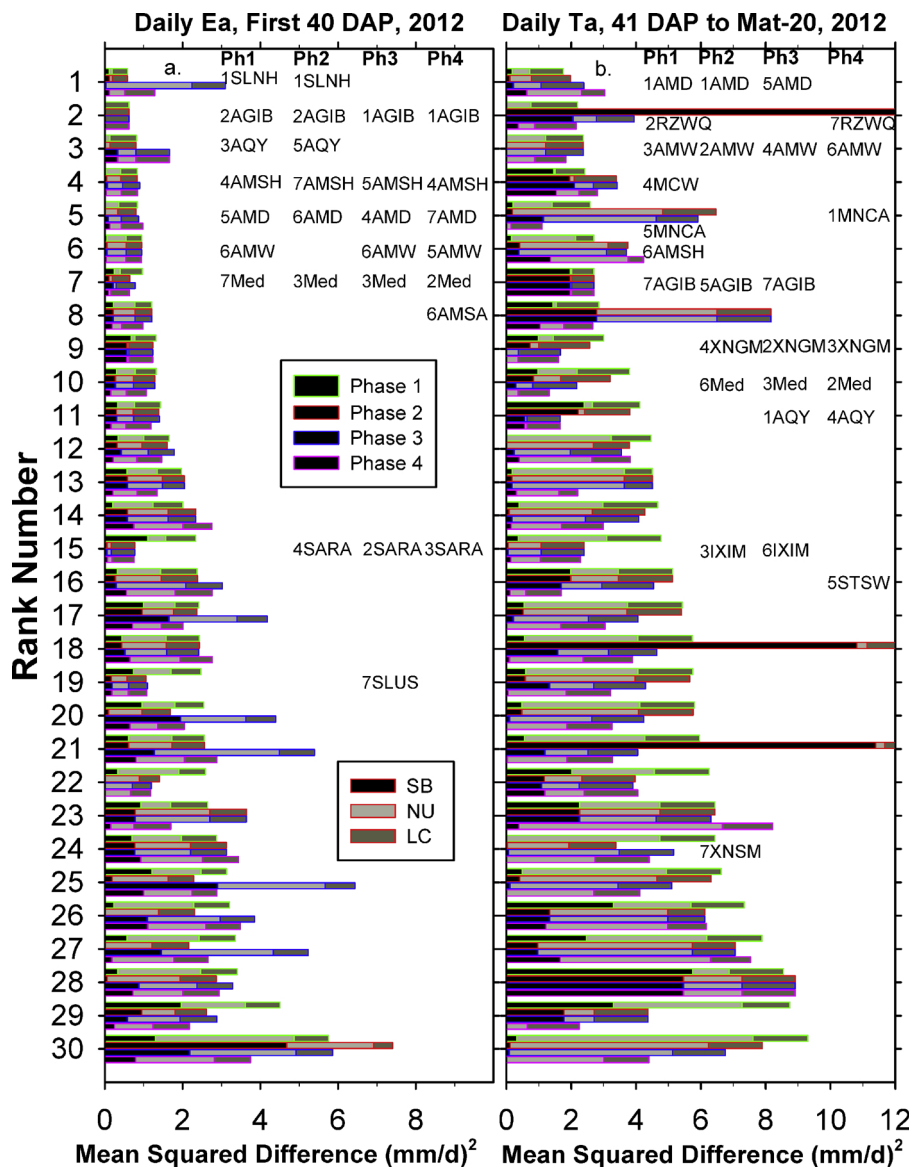


Fig. 4. Same as Fig. 3 except for “dry” 2012 growing season.

interesting. For example, focusing on the model for Rank No. 30 for the first 40 days in 2011, the model was quite far off from observations for Phases 1–3 in 2011 (Fig. 3a) but good in 2012 (Rank No. 6 in Fig. 4a), and then when all information was supplied in Phase 4, good agreement was achieved both years. The modeler for number 27 did far worse having LAI information in Phase 2 than in Phase 1 (Fig. 3a), yet ET predictions were improved in Phases 3 and 4. The median MSDs were relatively small for the first 40 days in both 2011 and 2012 (Figs. 3a, 4a), as well as mid-season in 2011 (Fig. 3b). However, there was steady improvement in the predictions with progressive phases in mid-season with the 2012 drought period for many of the models (Fig. 4b) except that the best ones for Phase 1 actually did worse as more information was supplied with successive phases.

The bar graphs in Figs. 3 and 4 additionally show the partitioning of the mean squared differences (MSD) from observations into standard bias (SB), non-unity of linearity (NU), and lack of correlation (LC). For the medians, most of the MSD is LC, and the amount of LC across the many models appears similar, which suggests this may be the experimental variability. Many of the models exhibit NU, and those with the highest MSD appear to be those with the greatest SB.

3.4. Comparisons of ET predictions for all eight years with observations

For the first 40 DAP, predictions of soil water loss (mostly E) did not change greatly with increasing information from Phases 1–4 (Fig. 5a). For half of the years, the models (and medians) tended to over-predict ET, whereas in 2008 and 2010, they under-predicted. Only in 2011 and 2013 are the box plots and medians close to the observations. These results suggest many of the models need improvements in simulating soil E and tend to over-estimate in dry periods as in 2012, although lower E attributed to unmeasured soil residue could also have been a variable factor.

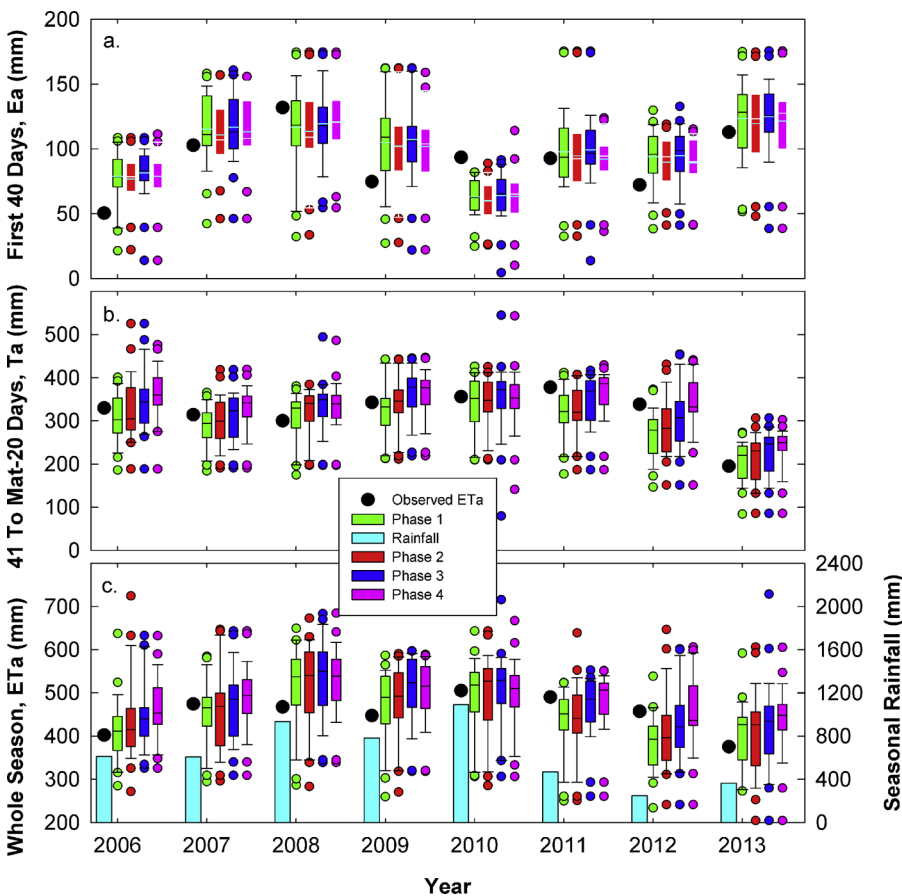


Fig. 5. (a.) Observed and simulated actual evapotranspiration for the first 40 days after sowing for eight years. During this period, the loss of water would be predominately actual soil evaporation, Ea. The box plots show the progressive phases for the simulated values for each year. (b.) Like (a.) but for the observed and simulated evapotranspiration from 41 days after sowing through 20 days before observed maturity. During this period the loss of water would be predominately actual transpiration, Ta. (c.) Like (b.) but for observed and simulated whole season-long evapotranspiration, ETa, from sowing through observed and simulated maturity dates, respectively. Also shown are the seasonal rainfall amounts for each of the eight years.

In contrast, during the mid-season 41 days after planting to 20 days before observed maturity, the box plots appear closer to the observations. There is growing improvement from Phase 1 to Phase 4, where the median is closer to observed and the range among the 25th to 75th percentile of the models represented by the boxes is shortened (Fig. 5b).

Besides predicting early season soil E and mid-season crop T well, accurate predictions of whole growing season ET (Fig. 5c) also require good predictions of the length of the growing season, i.e. maturity date (Fig. 6b). The modelers were given maturity dates for the last six years at the outset, so some parameterizations for phenology were done for Phase 1. Nevertheless, the median of the model predictions of seasonal ETa for the eight years did not get closer to the observed totals except for three of the years: 2007, 2011, and 2012 (Fig. 5c). The length of the box plots decreased going from Phase 1 to 4 for most of the eight years for season-long total actual ETa, which suggests the models tended to agree more and more with each other, but not with observations. The range from the lowest to the highest outliers was large, from about 300 to 600 mm, whereas the length of the boxes was less than 100 mm.

3.5. Comparison of predicted agronomic variables with observations

Maximum leaf area index (LAI) was underestimated by the median of the models in four of the years and was overestimated only in 2010, a wet year when water-logged soils caused reduced leaf area growth (Fig. 6a). Only in 2011 and 2013 were the medians close to observed LAI. Maturity dates for the last six years were supplied to the modelers at the outset, and as one would expect, there was no obvious improvement with additional information going from Phase 1 to 4 (Fig. 6b). However, there was a two-month range between the earliest and latest predictions of maturity date (Fig. 6b). Yet, the boxes themselves are fairly short, about 1 week, indicating close agreement among most of the models. In 2009 and 2010 the median was close to

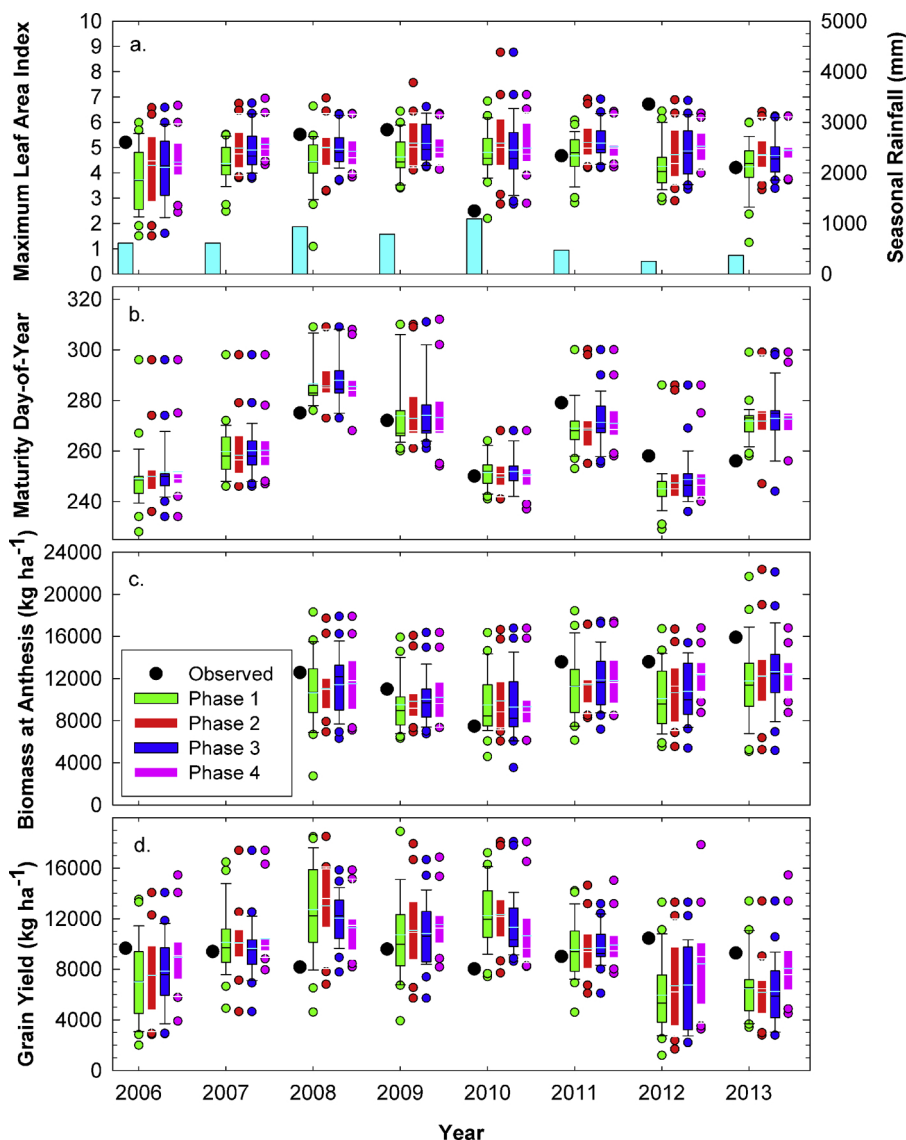


Fig. 6. (a.) Observed and simulated maximum leaf area indices for eight years. The box plots show the progressive phases for the simulated values for each year. Also shown are the seasonal rainfall amounts for each of the eight years. (b.) Like (a.) but for the observed and simulated maturity dates. (c.) Like (b.) but for observed and simulated above-ground biomass near anthesis. (d.) Like (b.) but for observed simulated grain yields.

observed, in 2008 and 2013 the median was about 10 days late, and in 2011 and 2012 it was about 10 days early, so overall agreement with observations was not good.

The median of the models generally underestimated biomass near anthesis except in the wet year 2010 when most of the models overestimated it (Fig. 6c). There was no obvious improvement with increasing information going from Phase 1 to 4. Predictions of grain yield varied by a factor of three from roughly 4000 to 12,000 kg/ha (Fig. 6d). The median was fairly close to observed in 2007, 2009, and 2011. However, the median underestimated in 2006, 2012, and 2013 and overestimated in 2008 and 2010. Thus, there is considerable room for improvement. The adjustments made by the modelers decreased the error and resulted in medians that were closer to observations for progressive phases as more information was supplied.

3.6. Performance of individual models in their ability to predict early E, mid-season T, and whole season ET for eight years

Several models had low mean squared differences (MSDs) that were close to the median for predictions of water loss (mostly E) during the first 40 days after planting for the eight years (Fig. 7a). Some modelers

made improvements going from one phase to another, but many did not. Three of the modeling groups (JUL, AMW, AMSA) made large improvements in Phase 4, placing them among the six best for this phase. For Phase 4 most of the MSD was due to lack of correlation (LC) for the median, and most of the models had similar LC. The models with the poorest performance had a large standard bias (SB).

Many of the models had MSDs close to or lower than the median for water loss from the 41st day after planting to 20 days before observed maturity (mostly T) for the eight years (Fig. 7b). For several of the models, there were no marked improvements with increasing information going from Phase 1 to 4 for this midseason part of the growing season with T dominant. Surprisingly, there were several that did worse in Phases 3 and 4, while at the same time several others (IXIM, AMW, STSW, AMSA, AQY) were among the best for Phase 4, but not for the first phases. Also like the first 40 days, the models with the poorest performance had a large standard bias (SB).

For total seasonal ETa for the eight years (Fig. 8a), about half of the models had MSDs comparable to that of the median, which was among the best for all four phases. Surprisingly, the best model (JUL) for Phase 1 and a few others did worse moving from Phase 1 to Phase 4. Like for midseason (Fig. 7b), several models (XNSM, MNCA, AMSA, AQY)

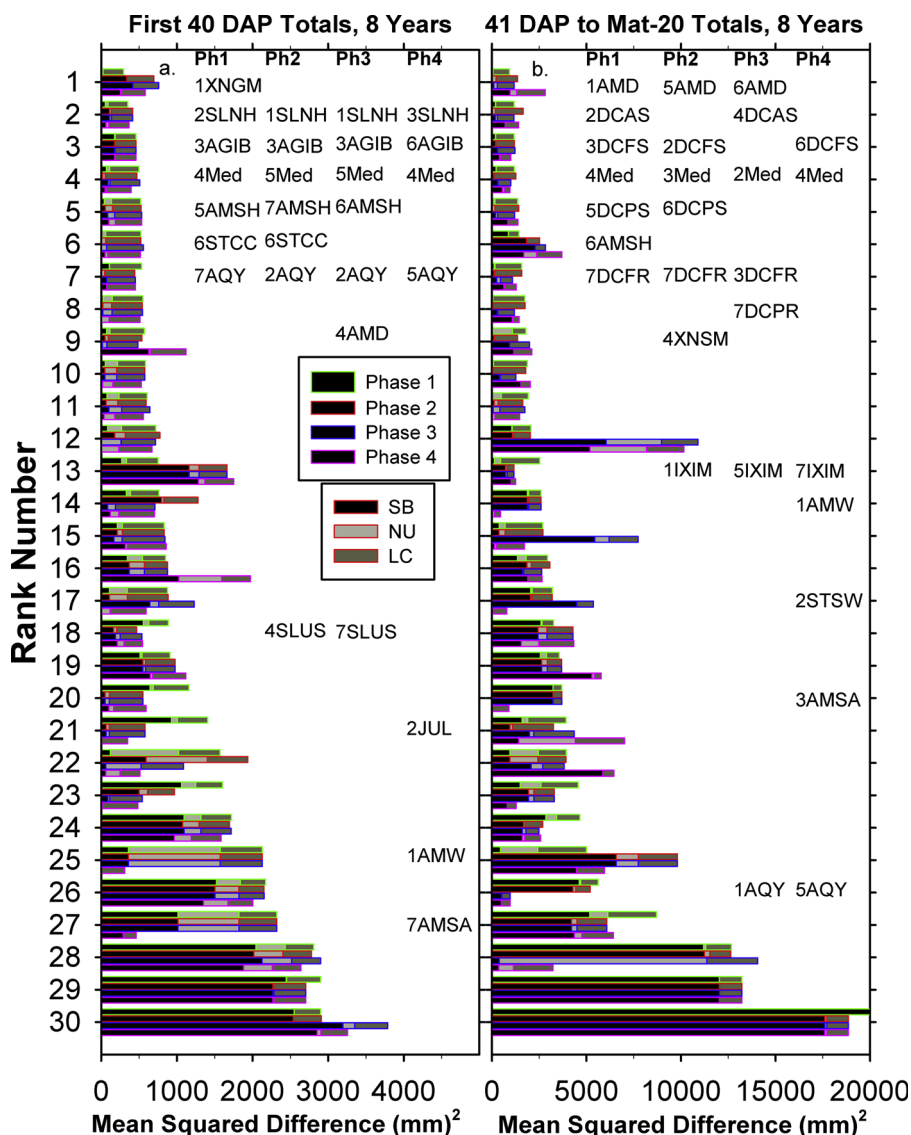


Fig. 7. (a.) Mean squared differences (MSD) between observed and simulated total water losses for the first 40 days after sowing for eight years for each of maize simulation models. Most water losses would have been from actual soil evaporation, E_a , with some transpiration from growing plant canopies during these periods. Notation like that of Fig. 3. (b.) Same as for (a.) except for total water losses from 41 days after sowing to 20 days before observed maturity for eight years. Most water loss would have been via actual transpiration, T_a , during these periods.

showed marked improvement for Phase 4. Again, large SBs were the main component of the large MSDs.

3.7. Performance of individual models in their ability to predict agronomic parameters for eight years

A few models (IXIM, SLFT, MCW, SLNH) had smaller MSDs for maximum leaf area index (LAI) for most phases than did the median (Fig. 9a). With increasing information from Phase 1 to 4, many of the modelers were able to decrease their LAI MSDs, and the median’s MSD also decreased. There was much variability among the models in their MSDs for biomass near anthesis (Fig. 9b). Like LAI, several of the models (SLFT, IXIM, RZWQ, DCPR, DCFR) had biomass MSDs that were smaller than that of the median for several of the phases. Three models (DCFS, DCAR, and DCPS) were greatly improved in Phase 4.

Likewise, there was much variability among the models for their grain yield MSDs (Fig. 8b), and several (MCW, AMSH, XNSM, SLNH, RZWQ, XNGM, AMSA, MNCA, AMW) were smaller than that for the median (Rank No. 10). Many of the modelers were able to lower their yield MSDs going from Phase 1 to 4, and the MSD for the median also

decreased. Note that the observed yield changed relatively little from year to year (Fig. 6d). Therefore, the best simulations resulted in graphs with a small cluster of points (Supplementary statistical data) with a small MSD even though the slope could vary wildly. Thus, for several of the models, non-unity of slope (NU) was the main component of the yield MSD, but for those with the largest MSDs and a couple others, standard bias (SB) was the major component (Fig. 8b).

4. Identification of “best” approaches, i.e. lowest mean squared differences from observations

Simulation of crop water use, growth, and yield mechanistically involves many processes. Early season soil water evaporation, E , and rate of leaf area development are important. If the initial estimates of E are too high, and the available soil water is depleted, then later season simulated ET will be too low because of a lack of water. Thus, there are many interactions and feedbacks, so it is difficult to identify specific approaches to the various processes as being “best.” Further, as the modelers were supplied more information going from Phase 1 to 4, different models emerged as being better than others.

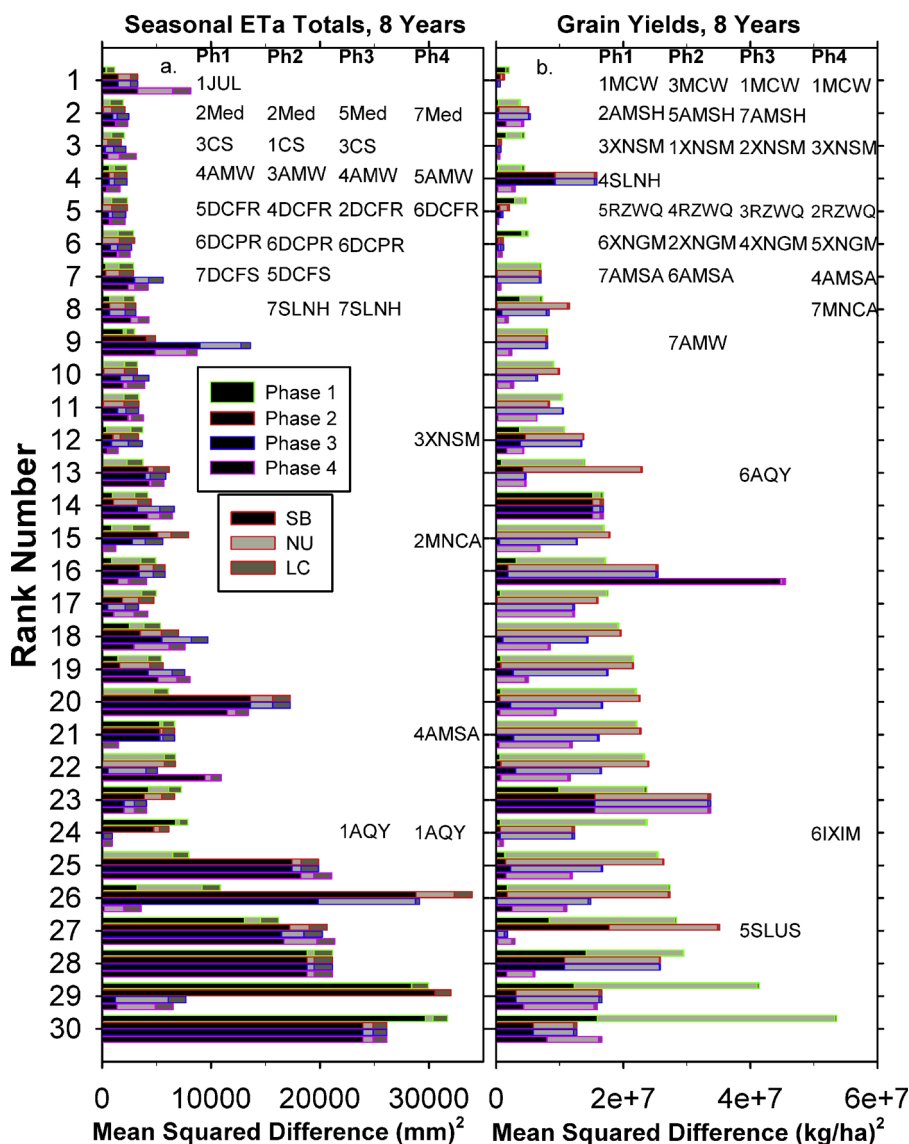


Fig. 8. (a). Similar to Fig. 7 except this is the MSDs for total seasonal (sowing through maturity) ETa values for the eight years of the experiment. Same notation as for Fig. 3. (b). Similar to (a) except that these are the MSDs for grain yields for all 8 years of the experiment.

4.1. Daily ETa (mostly E) for first 40 days after planting

Focusing on the first 40 days after planting when much of the water loss would have been due to soil E, the median proved best (lowest MSD) for 2011 for daily E for all four Phases, and it was among the top seven for daily ET for 2012 (Fig. 3a). The median was also among the top seven for total E for these first 40 day periods for all eight years (Fig. 7a). The SLNH model was among the best for this early period for daily E in 2011 (Fig. 3a) and 2012 (Fig. 4a) and also for the totals for all eight years (Fig. 7a). SLNH follows Hargreaves (1975) to estimate potential evaporation and dual crop coefficients to get actual Ea (Table S1). DCPR had a low MSD for the daily Ea for the first 40 days in 2011, but not in 2012 nor for the totals for the eight years. DCPR uses Priestly and Taylor (1972) to calculate potential ET, which is then partitioned into potential E and T based on leaf area. Actual E is calculated using constant and falling rate stages following Ritchie (1972). AMW had the lowest MSD for Phase 4 for the eight years (Fig. 7a). It uses Priestly and Taylor (1972) for potential soil E and then like DCPR follows Ritchie (1972) assuming constant and falling rate states for actual Ea (Table S1). AGIB is another model that did well predicting Ea for these initial

40 days for 2011 (Fig. 3a), 2012 (Fig. 4a), and for all eight years (Fig. 7a). AGIB uses a full iterative energy balance to solve for a surface temperature with separate soil and leaf fluxes (Table S1). Similarly, AMSH did well in 2011, 2012, and for all 8 years. AMSH, from the APSIM family of models, uses Priestly and Taylor (1972) to compute potential ET, from which it computes potential E. It estimates potential T from a potential dry matter increase times a transpiration efficiency. Actual T and E are calculated following Campbell (1985) accounting for soil water deficits. AQY did well for the eight-year seasonal totals for all phases, but was not generally among the best for the daily E in 2011 and 2012. AQY is a relatively simple model that does not simulate leaf area or biomass growth. E and T are functions of potential ET and a crop coefficient that varies with soil water availability (Table S1). The SARA modelers made good use of increasing information because it was second best for Phases 3 and 4 for 2011 (Fig. 3a), as well as 4th, 2nd and 3rd for Phases 2, 3, and 4, respectively, in 2012 (Fig. 4a). SARA follows FAO56 (Allen et al., 1998) with dual crop and soil coefficients as a function of LAI (Table S1). Thus, the best several models for simulating ET (mostly E) during these first 40 days after planting varied greatly in complexity and used a large variety of simulation approaches.

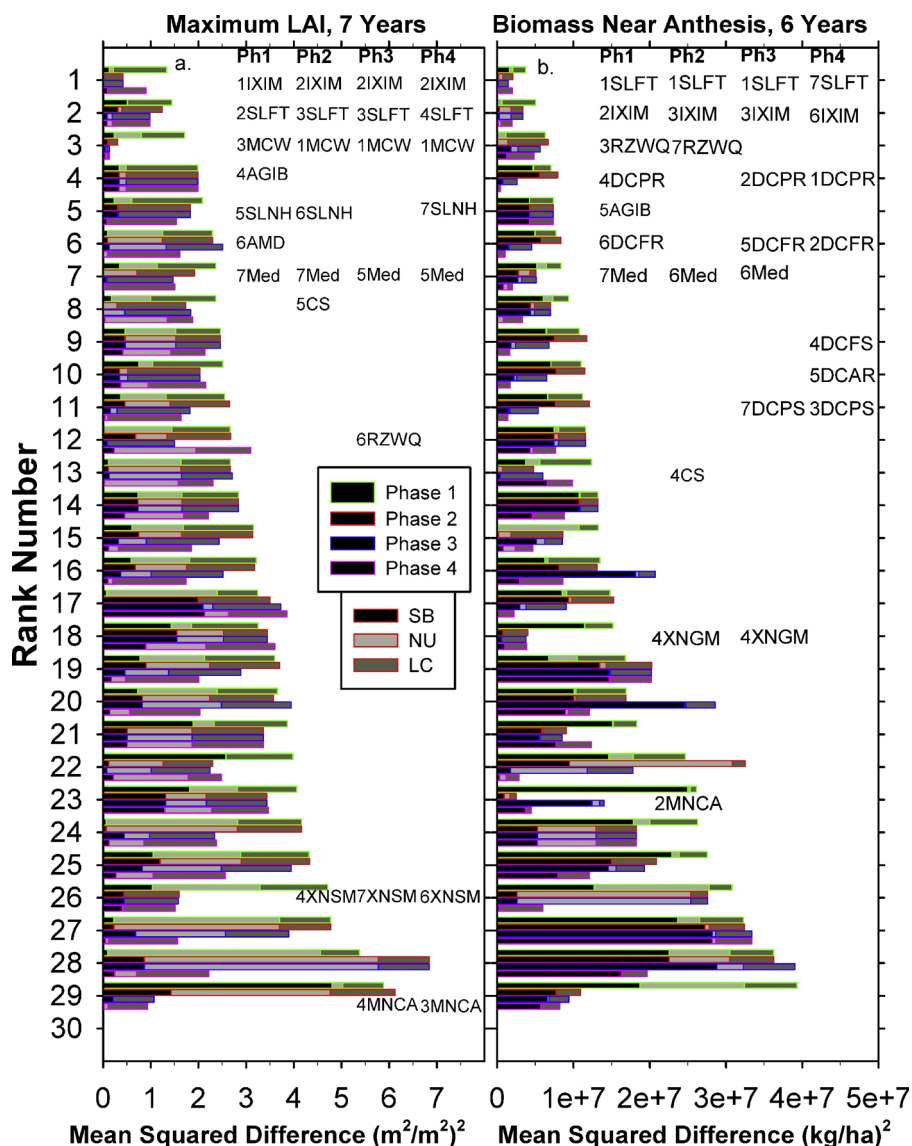


Fig. 9. (a). Mean squared differences (MSD) between observed and simulated maximum leaf area index, LAI, for seven years for each of maize simulation models. Notation like that of Fig. 3. (b.) Same as for (a.) except for above-ground biomass for six years near anthesis.

4.2. Daily *ETa* (mostly *T*) for mid-season from 41 days after planting to 20 days before observed maturity

Focusing next on the periods from 41 days after planting to 20 days before observed maturity, six models (XNSM, CS, SLFT, AMSH, DCFS, SNH) did better than the median for Phase 1 in 2011 (Fig. 3b), but the median was best for Phase 4 followed by SARA, DCFR, DCFS, AQY, SLFT and XNSM. Although XNSM did very well for daily *T* in 2011, but it was not among the best in 2012 (Fig. 4b) nor for the period totals for all eight years (Fig. 7b). XNSM uses Penman-Monteith FAO56 (Monteith, 1965; Allen et al., 1998) with a single crop coefficient (Table 2). Potential *E* and *T* are partitioned according to LAI. CS did well for Phases 1 and 2 in 2011, but was not among the best seven in 2012 or for the eight years. CS also uses FAO56 but the actual *T* is based on the potential gradient from soil to atmosphere and plant hydraulic properties (Table S1). DCFS was among the best for Phases 1, 3, and 4 in 2011 (Fig. 3b), but not 2012 (Fig. 4b), yet was among the best again for Phases 1, 2, and 4 for all eight years (Fig. 7b). DCFS follows FAO56 (Allen et al., 1998) to compute potential *ET* with potential *T* proportioned using leaf area and actual *T* then based on a soil-water-limited root water uptake (Table S1). IXIM also had lower MSDs for this mid-

season time period in 2011, 2012, and for all eight years, especially for Phases 2 and 3. IXIM uses Penman-Monteith for reference *ET* and partitions *E* and *T* based on LAI, with *Ea* based on Suleiman and Ritchie (2003) (Table S1). AQY also did well for Phases 3 and 4 during this mid-season period for 2011, 2012, and all eight years. As already mentioned, AQY is a simple model that computes *T* as a function of potential *ET* and a crop coefficient that varies with soil water availability (Table S1). After the Median, SARA had the second lowest MSD in 2011 for Phase 4. AMD was among the best models for Phases 1 through 3 in 2012 and the similar AMW for all four phases. They use a transpiration efficiency coefficient with potential crop growth to compute *T* (Table S1). RZWQ was among the best for Phases 1 and 4 for 2012. It uses the Shuttleworth and Wallace (1985) method for computing potential *ET* and then actual *T* is simulated by limiting root water uptake following Nimah and Hanks (1973). MNCA had the lowest MSD for Phase 4 for 2012. It also uses Penman-Monteith FAO56 for reference *ET* and multiplies a stage-specific crop coefficient to get actual (Table S1). *T* is based on water uptake by soil layer. Considering all eight years (Fig. 7b), besides AQY, DCFS, and IXIM already mentioned, AMW, STSW, and AMSA were among the best for Phase 4, while AMD, DCAS, DCPS, AMSH, and DCFR were among the best for Phase 1. AMW,

AMSA, AMD, and AMSH are all in the APSIM family that use a transpiration efficiency coefficient with potential crop growth to compute T (Table S1). DCAS, DCPS, and DCFR are all in the DSSAT family with varying ways to compute potential ET, soil water movement, and eventually T (Table S1). STSW used the Shuttleworth and Wallace (1985) method to compute potential ET followed by several steps to get E and T as per Brisson et al. (1998, 2003). Thus, like the first 40 days after planting, the best models for simulating ET (mostly T) during midseason varied in complexity and used a variety of approaches. However, we note there is relatively little overlap among the models identified as best for the first-40, E-dominated periods and the mid-season T-dominated periods [exceptions being AMW in 2012 (Fig. 4) and AQY for all eight years (Fig. 7)]. Thus, there ought to be room for model improvement by combining the better approaches for simulating E with those for simulating T.

4.3. Total ETa for whole growing seasons for eight years

Looking at the total seasonal ETa for all eight years (Fig. 8a), AQY had the lowest MSD for Phases 3 and 4. JUL did well for modeling ETa in the initial blind Phase 1, but was not among the seven best for the other phases, despite a significant improvement in LAI as a result of the phase 2 tuning. JUL is relatively complicated with an hourly time step and 10 LAI layers (Table S1). ET is calculated from the moisture potential gradient between soil and atmosphere times a resistance factor. CS, AMW, DCFR, DCPR, DCFs, XNSM, and MNCA all did well.

4.4. Grain yields for eight years

MCW had the lowest MSDs for grain yield for eight years for Phases 1, 3, and 4, and it was third for Phase 2 (Fig. 8b). MCW also uses Penman-Monteith FAO56 (Monteith, 1965; Allen et al., 1998) to calculate reference ET (Tao et al., 2009a). Evaporation was calculated from a 20-cm thick soil water evaporation layer following Ritchie et al. (1988). Transpiration is calculated soil-layer-wise and depends on soil water availability. XNSM also was among the best at simulating grain yield for all four phases, as well as mid-season Ta for all phases in 2011 (Fig. 3b) and whole season ETa for Phase 4. Its sister model, XNGM, also did well at predicting grain yield, but its name appears only once in the E, T, and ET tabulations of lowest MSD. AMSH and AMSA also had low MSDs for grain yield, and they were among the lowest seven for some of the phases of the E, T, and ET listings. RZWQ had a low MSD for grain yield, but considering water loss, it was among the best seven only for Phases 1 and 4 mid-season in 2012 (Fig. 4b).

4.5. Biomass and maximum leaf area index

SLFT had low MSDs for predicting above-ground biomass near anthesis and maximum leaf area (Fig. 9). Surprisingly, it was not among the best for grain yield, although it was among the best for mid-season ET in 2011 (Fig. 3b). Like several others, SLFT follows FAO56 (Monteith, 1965; Allen et al., 1998) for reference ET and partitions actual Ea and Ta using dual crop coefficients. Both water availability and energy availability are used to limit water, and heat stress effects are calculated using canopy temperature. DCPR and several other of the DC_ (DSSAT-CSM family) models predicted biomass near anthesis well, especially Phase 4, but were not among the best seven for LAI or grain yield. IXIM, AGIB, CS, MNCA, and RZWQ also had low MSDs for biomass for various phases (Fig. 9b). IXIM, MCW, AGIB, SLNH, AMD, CS, XNSM, and MNCA all had low MSDs for the prediction of maximum LAI (Fig. 9a).

4.6. Overall rankings considering both ET and agronomic factors

Overall rankings of the best seven models considering (1) ETa for the first 40 days after planting (mostly Ea), (2) ETa for the 41st through

20 days before maturity (mostly Ta), (3) ETa for entire growing seasons, (4) maximum leaf area index, (5) biomass near anthesis, and (6) grain yield for all eight years (seven years for LAI and six for biomass) are shown in Table 3. These rankings were calculated using the rankings of the MSDs for all the models (Figs. 3, 4 and 7–9, Supplementary statistical data) for the six parameters. For this overall perspective of both water use and agronomic performance of the models, the medians were the best for all four phases. Thus, for this overall perspective, a multi-model approach would be best for doing predictions rather than a particular individual model.

However, an examination of individual models reveals that SLNH did well for the first and second phases, but was not among the best seven for Phases 3 or 4. Except for DCPR in Phase 2, DCFR and DCPR were among the best for all phases. CS and IXIM were among the seven best for the first three phases. XNGM and XNSM were among the seven best for Phases 1 and 2. AQY appears next best the median for Phase 3 and in 3rd place for Phase 4. However, AQY is problematic for this inter-comparison because it does not compute LAI or biomass, so its ranking for yield was additionally used for LAI and biomass. For Phase 4, AMW and AMSA ranked among the best seven, and MNCA also performed well enough to appear in Table 3 for Phase 4.

5. Inter-comparisons among models within the same family

In several cases within this study modelers used models from within the same family but changed only a particular aspect. Looking at these cases in detail allows identification of which approaches are better compared to others while most features of the plant growth and water use simulations are kept constant.

5.1. AMW vs AMSA vs AMD vs AMSH (Table 4)

In the case of the APSIM model, two modelers independently used the APSIM-SWIM (AMSH, AMSA; soil water balance is based on Richard’s equation) and APSIM-SoilWat (AMD, AMW; soil water balance is based on a tipping bucket approach) models (Table 4). Overall, the APSIM models performed relatively well in comparison to the other models across the four phases (Figs. 3, 4 and 7–9, Table 3). Interestingly, we did not find that the more complex approach based on the Richard’s equation (SWIM, Huth et al., 2012) outperformed the standard tipping bucket approach in APSIM (Probert et al., 1998) across all years. For instance, LAI was most accurately simulated by AMD and AMW (two different users). Also for ET during the mid-season, there was a slight trend that the tipping bucket approach performed better (Table 4). However, in terms of grain yield prediction, the APSIM-SWIM (AMSA user) substantially outperformed all other APSIM models (Phase 4, RMSD of 871 kg/ha versus 1916 kg/ha). In general, the results of two different calibration approaches can be observed in the data. For AMD

Table 3

Overall ranking of the best seven models considering ETa for the first 40 days after planting (mostly Ea), ETa for the 41st through 20 days before maturity (mostly Ta), ETa for entire growing seasons, maximum leaf area index, biomass near anthesis, and grain yield for all eight years (six years for LAI and biomass) presented separately for each of the four phases of the study. These rankings were calculated using the rankings of the MSDs for all the models for the six parameters and the four phases. (Note that the AQY model does not compute LAI or biomass, so for this model, its ranking for yield was additionally used for LAI and biomass.)

Phase	Best	2nd	3rd	4th	5th	6th	7th
1	Median	SLNH	DCPR	CS	DCFR	XNGM	IXIM
2	Median	CS	IXIM	XNSM	SLNH	XNGM	DCFR
3	Median	AQY	SLFT	DCPR	DCFR	IXIM	CS
4	Median	AMW	AQY	DCFR	AMSA	DCPR	MNCA

Table 4

Root Mean Square Differences among models of the same APSIM and MAZSIM families, but with changes in particular aspects. APSIM-A and -H indicate different users.

Root Mean Square Differences	Phase	APSIM Soil Wat AMW	APSIM SWIM- A AMSA	APSIM Default Soil Wat AMD	APSIM SWIM-H AMSH	MAZSIM daily MZD	MAZSIM hourly MZH
Daily water loss during first 40 days after sowing in 2011 (Ea, mm/d)	1	2.97	1.37	1.20	1.14	1.46	1.49
	2	2.97	1.37	1.19	1.12	1.59	1.59
	3	2.97	1.37	1.08	1.11	1.59	1.59
	4	1.11	1.27	1.38	1.11	1.59	1.60
41 days after sowing to maturity-20 days in 2011 (Ta, mm/d)	1	1.44	2.93	1.32	1.26	1.37	1.47
	2	1.44	2.93	1.33	1.33	1.91	1.91
	3	1.44	2.93	1.31	1.35	1.91	1.91
	4	1.28	1.27	1.40	1.28	1.50	1.89
Daily water loss during first 40 days after sowing in 2012 (Ea, mm/d)	1	0.97	1.10	0.92	0.92	1.42	1.70
	2	0.97	1.10	0.90	0.91	1.53	1.77
	3	0.97	1.10	0.93	0.95	1.53	1.77
	4	0.97	0.99	0.99	0.92	1.66	1.86
41 days after sowing to maturity-20 days in 2012 (Ta, mm/d)	1	1.54	2.12	1.33	1.64	2.71	2.81
	2	1.54	2.12	1.41	1.94	2.47	2.66
	3	1.54	2.12	1.55	1.93	2.47	2.66
	4	1.36	1.48	1.74	2.06	2.48	2.74
Total water loss during first 40 days after sowing all 8 years (Ea, mm)	1	46.2	48.2	23.9	22.8	27.4	30.2
	2	46.2	48.2	23.2	23.0	40.8	31.3
	3	46.2	48.2	22.0	23.1	40.8	31.3
	4	17.6	21.5	33.5	23.1	41.9	33.5
41 days after sowing to maturity-20 days all 8 years (Ta, mm)	1	51.0	60.8	30.5	37.9	93.3	70.7
	2	51.0	60.8	36.8	50.2	78.0	99.0
	3	51.0	60.8	34.3	53.2	78.0	99.0
	4	21.3	30.1	53.1	60.9	80.1	77.1
Total seasonal water loss for 8 years (ETa, mm)	1	48.2	81.6	73.3	127.3	89.2	78.1
	2	48.2	81.6	83.8	143.7	141.0	131.5
	3	48.2	81.6	98.6	142.2	141.0	131.5
	4	40.6	39.0	87.3	146.2	145.2	116.0
Grain Yields for 8 years (kg/ha)	1	2848	2666	3287	1958	7320	6435
	2	2848	2666	3715	2263	3561	4080
	3	2848	2666	3674	2314	3561	4080
	4	1581	871	2087	2081	4079	3991
Above-ground biomass for 6 years (kg/ha)	1	3406	3633	5248	4281	5684	4101
	2	3406	3633	4571	3016	5693	4503
	3	3406	3633	4396	2919	5780	4499
	4	2772	2975	3478	3515	5780	4499
Maximum leaf area index for 7 years (LAI, m ² /m ²)	1	1.57	1.69	1.51	1.64	2.02	1.80
	2	1.57	1.69	1.51	1.64	1.85	1.86
	3	1.57	1.69	1.58	1.65	1.85	1.86
	4	1.46	1.49	1.27	1.52	1.86	1.90

and AMSH, one modeler aimed to take into account the new information provided at each calibration step and altered model parameters to match predictions with the new observed information. Soil parameters [runoff number, crop-specific soil water extraction factor (called KI in APSIM), in case of AMD also the first and second stage evaporation term called cona and u, and at Phase 4 for AMSH, a water table was introduced] were modified from phase to phase, with a focus shifting towards the Phase 4 calibration step to improve grain yield and LAI prediction. Interestingly, the yield and LAI improvements came with a reduced prediction accuracy of the water dynamics.

The second modeler (AMSA, AMW) initialized cultivar and the soil parameters in Phase 1 but did not make further changes until final Phase 4, when all observed information was made available. During calibration, this modeler aimed to better represent the overall system and to adjust only those model parameters that are known to characterize a particular site. First, the water table option was activated in APSIM-SWIM to better reflect the local conditions of this site that is known to have shallow water tables (Logsdon et al., 2009; Fan et al., 2013; Ordóñez et al., 2018; Rizzo et al., 2018). The water table inhibited root growth (by activating also the air-filled pore space parameter). The simulated maximum root depth was shallow in wet years

(2008 and 2010), deep in dry years (2012) and around average root depth in normal years (about 1.5 m for this area; Ordóñez et al., 2018). This combined change in the APSIM-SWIM model improved year by year predictability of yield. In APSIM-SoilWat there is no mechanistic option to simulate water table dynamics, so it was not considered. Additional changes made to both APSIM SWIM and SoilWat models were: 1) decrease the potential kernel number and potential kernel growth rate (cultivar-specific parameters) to improve yield predictions; 2) decreased the “svp_fract” parameter from 0.75 to 0.62 to improve simulation of daily transpiration rate. This parameter is used in the calculation of vapor pressure deficit that drives transpiration and its value depends on local relative humidity.

In summary, without in-depth knowledge of the site conditions (Phase 1), APSIM performed among the best models and continued to do so as more information was supplied in Phases 2 through 4 (Figs. 3, 4 and 7–9, Table 3). The question of whether SWIM or default tipping bucket approach is superior depends on the research objective. If the objective is an accurate simulation of grain yield in this area, the SWIM version with activated water tables and inhibition to root depth seems a good option. For simulation of ET, the SoilWat model performed slightly better than SWIM. However, more data for testing, in

particular, soil water measurements coupled with subsoil water table dynamics, ET measurements and soil cover would be necessary.

5.2. MZD vs MZH (Table 4)

Both the daily and hourly versions of MAZSIM (MZD and MZH) operate on an hourly time step for plant processes, including solving an energy balance (Table 1). The soil model uses a finite element formulation and the time step can vary from seconds to hours depending on when input data (boundary condition) are read and convergence criteria. The daily version of the model (MZD) has the same code for soil and plant processes as the hourly version (MZH), but both versions require hourly weather data. For MZH, the hourly data are input directly. For MZD, hourly data are generated from daily observations. These daily data include maximum and minimum temperatures, total precipitation and solar radiation, and average wind speed and relative humidity. Precipitation is distributed for each hour where the maximum precipitation for that hour is no larger than the average hourly rainfall rate for that location. The daily temperature and solar radiation values are interpolated to an hourly basis (Timlin et al., 2002) and the model executes similarly as the hourly version.

The RMS differences in Table 4 between the hourly MZH and daily MZD versions are mainly due to differences in how an hourly time step model such as MAZSIM distributes temperature and precipitation over a 24 h period when only daily values are available. For example, hourly rainfall data contains rainfall data to the nearest hour. Thus, precipitation is simulated close to the actual time precipitation occurred. When only daily data are available, MZD assumes rainfall starts at midnight. The totally daily rainfall is distributed hourly using the average hourly rainfall rate. If there is 12 mm of precipitation and average precipitation rate for the site is 3 mm per hour, then precipitation is distributed in 3 mm pieces over a four-hour time period after midnight. Thus, precipitation always occurs at night. MZD assumes maximum temperature occurs at midday which may be incorrect for some time periods, especially if cloud cover follows or precedes midday. Weather stations do not generally record the time of the maximum temperature, so the distribution of simulated hourly temperatures can differ from observed. This would affect simulated photosynthesis and ET values as the simulated temperature/light values for a particular hour could be quite different from the observed pairs. In terms of simulated results, there did not seem to be consistent patterns in the errors in Table 4 for ET, biomass or yield that would indicate that one approach was better than the other.

5.3. DCAS vs DCFS vs DCPS vs DCAR vs DCFR vs DCPR (Table 5)

Within the DSSAT models, the DCAR, DCFR, and DCPR methods, which used the Ritchie (1972) method (constant rate stage one followed by square root of time falling rate stage two) to calculate soil water evaporation, often outperformed a newer approach by Suleiman and Ritchie (2003) and Ritchie et al. (2009). In particular, for early season ET in 2011, 2012, and over all 8 years, MSDs were generally lower for DCAR, DCFR, and DCPR compared to DCAS, DCFS, and DCPS (Table 5). Similarly, the MSDs for total seasonal water loss over all 8 years, grain yield, biomass, and LAI were also lower. However, the Suleiman and Ritchie (2003) and Ritchie et al. (2009) soil water evaporation approach outperformed the Ritchie (1972) approach for ET data from 41 days after sowing to maturity over all 8 years. Although the Ritchie (1972) method performed well in this study, it limits soil water evaporation to water contents in the top soil layer only, whereas the Suleiman and Ritchie (2003) approach calculates up-flux from the deeper soil layers. Future work should improve the Suleiman and Ritchie (2003) method by adjusting the currently hard-coded coefficients that influence the algorithm performance and testing against high-quality soil water evaporation data.

For simulation of yield, biomass, and LAI, the study demonstrated

improved results using the Priestly and Taylor (1972) ET approaches (DCPS and DCPR). Also, the Priestley-Taylor approaches best calculated ET during the drought year of 2012. The result was unexpected because the Priestley-Taylor approaches do not consider wind speed nor dew point temperature for ET calculations, which should likely be necessary for accurate simulation during the water-limited 2012 growing season. In this study, the native DSSAT ET methods (DCFS, DCPS, DCFR, and DCPR) performed better than the ASCE dual Kc approach, which is formulated from the most recent ET standardization (Allen et al., 2005). Future work should further compare these ET methods using data sets from an arid region, where water supply and ET can be more carefully controlled via irrigation and where high precipitation and subsurface drainage systems do not confound the water balances of the cropping system.

5.4. SLNH vs SLFT (Table 6)

Considering the SIMPLACE-LINTUL5 models, for the estimation of ETp, Hargreaves (1975) and FAO56 Penman-Monteith (Monteith, 1965; Allen et al., 1998) approaches were used in SLNH and SLFT, respectively. The evaporation losses (Ea) in the early crop growth stages (first 40 days after sowing) were better captured by SLNH in the individual years 2011, 2012 and also across 8 years compared to the values estimated by SLFT (Table 6). In contrast, transpiration losses (Ta) in the year 2011 were better estimated by SLFT, but in the drier year (i.e., 2012), SLNH performed better. Across 8 years of simulation, SLFT estimated total seasonal water losses (ETa) better than SLNH. These differences could be attributed to a more accurate estimation of ETp by Penman-Monteith especially when the air temperature is high (later part of the growing season or during the dry year 2012).

Simulated above-ground biomass was better estimated by SLFT which could be related to the better estimation of LAI compared to the estimation by SLNH. The two models differed in the parameterization of LAI, with SLNH having more rapid initial LAI and higher specific leaf area, which is compensated by using low values of radiation use efficiency. On the other hand, SLFT had a much lower initial rate of leaf area expansion and lower specific leaf area, compensated by a higher RUE, more typical of maize. In contrast, simulated yield values were better estimated by SLNH, which suggests the harvest index and senescence aspects of SLFT can be improved.

5.5. STCC vs STSW (Table 6)

Within the STICS family, STCC and STSW use identical routines excepting for calculating daily values of crop water demand and soil water evaporation (Table S1). STCC is a classical “crop coefficient” approach using the potential evaporation ETp as a driving variable (Brisson et al., 1992). Crop water demand and soil water evaporation use the same process of calculation in two steps. Firstly, potential fluxes related to energy available and secondly actual fluxes related to water availability are calculated. STSW is an alternative approach in which determinations of the two fluxes use the Shuttleworth and Wallace (1985) daily time step model. The evaporation from the soil and the crop water demand are written under the form of Penman-Monteith combination equation using a resistance to diffusion between canopy and soil cover (RAS) and another aerodynamic resistance to a reference level above the canopy (RAA) (Brisson et al., 1998).

In both cases, the water balance is used to calculate the water status of the soil and the crop as well as the water stress indices that reduce leaf growth and net photosynthesis. Soil water evaporation is distributed over the soil profile based on the exponential decline of the contribution of each 1 cm soil layer. Actual crop transpiration is the minimum of soil water supply and crop water demand. Water uptake of each 1 cm soil layer is driven by root growth, root density and soil water status. Water uptake is possible if the soil water content is greater than the water content at wilting point.

Table 5
 Root Mean Square Differences among models of the same DSSAT-CERES Maize family, but with changes in particular aspects.

Root Mean Square Differences		ASCE Dual Kc		FAO56 EORATIO		Priestley-Taylor		ASCE Dual Kc		FAO56 EORATIO		Priestley-Taylor	
		Suleiman DCAS		Suleiman DCFS		Suleiman DCPS		Richie DCAR		Richie DCFR		Richie DCFR	
Phase													
Daily water loss during first 40 days after sowing in 2011 (Ea, mm)	1	1.80	1.50	1.69	1.16	2.04	1.09	1.16	2.40	1.09	1.09	1.16	1.09
	2	1.80	1.50	1.69	1.16	2.02	1.09	1.16	2.02	1.09	1.09	1.16	1.09
	3	1.81	1.50	1.69	2.07	2.03	1.11	2.07	2.03	1.09	1.11	2.07	1.09
	4	1.79	1.50	1.68	1.17	1.09	1.11	1.17	1.09	1.09	1.11	1.17	1.11
41 days after sowing to maturity-20 days in 2011 (Ta, mm)	1	1.71	1.28	1.48	1.71	2.39	1.62	1.71	2.39	1.62	1.62	1.71	1.62
	2	1.73	1.32	1.54	1.70	2.39	1.62	1.70	2.39	1.62	1.62	1.70	1.62
	3	1.12	0.98	1.10	2.29	2.21	1.20	2.29	2.21	1.20	1.20	2.29	1.20
	4	1.04	0.95	1.07	1.04	0.94	1.08	1.04	0.94	1.08	1.08	1.04	1.08
Daily water loss during first 40 days after sowing in 2012 (Ea, mm)	1	1.77	1.60	1.55	2.40	2.44	2.40	2.40	2.44	2.40	2.40	2.40	2.40
	2	1.51	1.30	1.54	1.27	1.18	1.13	1.27	1.18	1.13	1.13	1.27	1.13
	3	2.54	2.10	2.04	1.34	1.19	1.13	1.34	1.19	1.13	1.13	1.34	1.13
	4	1.70	1.43	1.42	1.21	1.09	1.04	1.21	1.09	1.04	1.04	1.21	1.04
41 days after sowing to maturity-20 days in 2012 (Ta, mm)	1	1.29	1.20	1.15	2.57	2.41	2.33	2.57	2.41	2.33	2.33	2.57	2.33
	2	3.58	3.65	2.38	2.51	2.40	2.32	2.51	2.40	2.32	2.32	2.51	2.32
	3	2.16	2.01	2.07	2.26	2.06	2.01	2.26	2.06	2.01	2.01	2.26	2.01
	4	1.97	1.81	1.79	2.03	1.81	1.75	2.03	1.81	1.75	1.75	2.03	1.75
Total water loss during first 40 days after sowing all 8 years (Ea, mm)	1	53.0	41.5	46.6	24.6	23.5	24.2	24.6	23.5	24.2	24.2	24.6	24.2
	2	52.7	41.2	46.4	24.4	23.3	24.0	24.4	23.3	24.0	24.0	24.4	24.0
	3	53.8	41.5	46.4	25.4	23.4	24.0	25.4	23.4	24.0	24.0	25.4	24.0
	4	51.4	39.8	44.8	23.6	22.7	23.0	23.6	22.7	23.0	23.0	23.6	23.0
41 days after sowing to maturity-20 days all 8 years (Ta, mm)	1	34.3	34.3	37.0	43.3	39.7	41.8	43.3	39.7	41.8	41.8	43.3	41.8
	2	40.6	34.4	37.7	42.3	39.7	41.8	42.3	39.7	41.8	41.8	42.3	41.8
	3	34.1	34.8	34.6	35.5	33.0	34.5	35.5	33.0	34.5	34.5	35.5	34.5
	4	37.6	31.6	37.1	45.1	35.9	38.0	45.1	35.9	38.0	38.0	45.1	38.0
Total seasonal water loss for 8 years (ETa, mm)	1	73.8	53.9	64.9	58.1	48.3	53.8	58.1	48.3	53.8	53.8	58.1	53.8
	2	74.9	53.7	67.2	58.3	48.3	54.6	58.3	48.3	54.6	54.6	58.3	54.6
	3	87.2	74.8	81.4	58.1	46.6	52.1	58.1	46.6	52.1	52.1	58.1	52.1
	4	89.9	65.1	80.4	61.6	46.2	50.6	61.6	46.2	50.6	50.6	61.6	50.6
Grain Yields for 8 years (kg/ha)	1	5045	4709	4833	4696	4397	4126	4696	4397	4126	4126	4696	4126
	2	5135	4774	4902	4761	4441	4229	4761	4441	4229	4229	4761	4229
	3	4096	4026	4075	4096	3805	3579	4096	3805	3579	3579	4096	3579
	4	3461	3462	3418	3070	2920	2626	3070	2920	2626	2626	3070	2626
Above-ground biomass for 6 years (kg/ha)	1	3283	3350	5684	3320	2781	2662	3320	2781	2662	2662	3320	2662
	2	3911	3433	3475	3392	2896	2819	3392	2896	2819	2819	3392	2819
	3	3006	2606	2326	2554	2132	1621	2554	2132	1621	1621	2554	1621
	4	1481	1281	1179	1312	1027	686	1312	1027	686	686	1312	686
Maximum leaf area index for 7 years (LAI, m ² /m ²)	1	2.08	1.90	1.78	1.91	1.79	1.60	1.91	1.79	1.60	1.60	1.91	1.60
	2	2.08	1.92	1.77	1.89	1.78	1.63	1.89	1.78	1.63	1.63	1.89	1.63
	3	1.99	1.70	1.56	1.99	1.59	1.35	1.99	1.59	1.35	1.35	1.99	1.35
	4	1.60	1.42	1.36	1.43	1.32	1.28	1.43	1.32	1.28	1.28	1.43	1.28

Table 6
 Root Mean Square Differences among models of the same SIMPLACE-LINTUL5, STICS-ETP, and Expert-N families but differ in particular aspects.

Root Mean Square Differences		Phase	SIM-PLACE LINTUL5 Hargreaves ET _o SLNH	SIM-PLACE LINTUL5 FAO56 Canopy T SLFT	STICS ETP READ Penman	STICS ETP SW calc	Expert-N GECROS	Expert-N SPASS
					STCC	STSW	XNGM	XNSM
Daily water loss during first 40 days after sowing in 2011 (Ea, mm)		1	1.05	1.67	1.18	1.45	1.61	1.69
		2	1.06	1.42	1.18	1.45	1.72	1.70
		3	1.65	1.32	1.18	1.45	1.49	1.46
		4	1.05	1.25	1.21	1.39	1.40	1.35
41 days after sowing to maturity-20 days in 2011 (Ta, mm)		1	1.30	1.22	1.82	1.55	1.46	0.93
		2	1.31	1.16	1.82	1.55	1.51	1.01
		3	2.08	1.01	1.81	1.60	1.63	0.92
		4	1.22	0.98	1.36	1.20	1.57	0.98
Daily water loss during first 40 days after sowing in 2012 (Ea, mm)		1	0.77	2.12	1.56	1.60	1.85	1.79
		2	0.76	1.62	1.56	1.60	1.69	1.52
		3	1.76	1.70	1.55	2.32	1.82	1.96
		4	1.14	1.48	1.66	1.70	1.66	1.70
41 days after sowing to maturity-20 days in 2012 (Ta, mm)		1	2.11	3.05	2.54	2.27	1.73	2.54
		2	1.95	2.81	2.54	2.27	1.60	1.84
		3	1.88	2.60	2.51	2.13	1.29	2.27
		4	1.96	2.10	2.87	1.30	1.27	2.10
Total water loss during first 40 days after sowing all 8 years (Ea, mm)		1	18.8	40.1	22.9	28.8	17.2	27.6
		2	20.4	31.1	22.9	28.8	26.4	35.8
		3	20.4	23.3	23.6	29.1	27.6	26.6
		4	19.2	22.0	22.7	29.3	24.1	26.5
41 days after sowing to maturity-20 days all 8 years (Ta, mm)		1	59.7	54.2	62.7	56.6	45.4	42.4
		2	60.7	55.4	62.6	56.5	45.1	36.7
		3	60.7	51.3	61.8	73.3	104.4	44.6
		4	76.0	51.5	80.3	28.1	100.7	45.9
Total seasonal water loss for 8 years (ETa, mm)		1	54.5	85.2	82.1	56.9	54.6	61.1
		2	55.8	81.5	82.1	56.9	70.1	57.7
		3	55.8	63.8	71.3	65.6	116.7	60.9
		4	65.6	63.8	104.7	62.6	93.5	38.5
Grain Yields for 8 years (kg/ha)		1	2109	4200	4651	5231	2248	2108
		2	3985	4009	4651	5231	1067	913
		3	3985	3510	4195	3858	1090	873
		4	1727	3510	2250	3343	972	785
Above-ground biomass for 6 years (kg/ha)		1	3061	1935	6021	4111	3903	6272
		2	2647	1427	6021	4111	2005	3309
		3	2647	1209	6250	5346	1944	3068
		4	1819	1410	4437	3476	1966	2869
Maximum leaf area index for 7 years (LA1, m ² /m ²)		1	1.44	1.20	2.19	2.04	1.80	2.17
		2	1.35	1.12	2.19	2.04	1.87	1.26
		3	1.35	0.99	1.97	1.53	1.93	1.26
		4	1.24	0.99	1.25	1.54	1.97	1.23

An important difference between STCC and STSW is that STSW can be used to take into account the effects of increasing of atmospheric CO₂, whereas STCC cannot, so the pertinent differences between the models for this study is their approaches for computing potential ET. Because CO₂ was not a factor in this study, large differences between the two models were not expected, and indeed such was the case (Table 6). The magnitudes of the calculated daily ET over the periods of simulation are comparable with the measured values. The best estimates were obtained for the years 2007, 2009, 2010, 2011, and the worst for the years 2006, 2008, 2012 and 2013. Based on a statistical analysis of the comparison between the observed and calculated ET over the 8 years, STCC gave slightly better estimates than STSW, and STCC tended to do a little better at simulating E during the first 40 days after planting, whereas STSW did a little better simulating T during mid-season. STSW was better at simulating biomass, whereas STCC was better at simulating grain yield. Thus, these results raise doubts about the quality of the parameterization of the LAI computations (plant parameters) and the relevance of the formalisms which simulate the water fluxes (soil evaporation and transpiration) that depend on LAI.

5.6. XNGM vs XNSM (Table 6)

Within the Expert-N family, XNGM and XNSM use identical routines to simulate soil moisture dynamics (Richards equation following Simunek et al., 1998), solute and soil heat transport (Hutson and Wagenet, 1992), and soil organic matter turnover (Johnsson et al., 1987), but have their own sub-modules for simulating crop growth and evapotranspiration (Table S1). In XNSM, potential evapotranspiration ET_p is simulated following Penman-Monteith FAO56 (Monteith, 1965; Allen et al., 1998) using air temperature and a single crop coefficient which depends on development stage. Partitioning between potential transpiration and soil water evaporation is calculated from the leaf area index, assuming no soil water evaporation when LAI exceeds a value of 3. XNGM also simulates ET_p using the Penman-Monteith approach but calculates leaf and soil surface temperatures by closing the energy balance. In addition, boundary layer resistances and turbulent resistances for water and heat are calculated for the canopy and the soil surface layer depending on crop height and wind speed, which are directly used to calculate T_p and E_p. In XNGM, actual soil water evaporation E is limited by soil water availability in the uppermost soil layer (5 cm), whereas in XNSM, E is limited by maximum flux rate from the soil surface. In both models, actual transpiration is simulated by summing up soil-layer-wise root water uptake, but have different strategies to represent root system formation. XNGM assumes a uniform distribution of root length density within the rooted zone, where rooting depth increases with increasing root biomass, neglecting any

interaction with soil temperature and resource availability. In contrast, XNSM simulates root length distribution and vertical extension of the root system by taking into account the potential impact of unfavorable environmental factors in the rooted soil layer. In this model, besides soil water availability, also a maximum root water uptake capacity per unit root length limits root water uptake from the single soil layers.

The different approaches pursued in the two models led to very similar model accuracies when simulating E during the soil dominated first 40 days after sowing, including the two contrasting years 2011 and 2012 (Table 6). During this early phase, differences in simulated root system played only a minor role, and both methods to simulate E_p had similar results as simulated LAI and plant height are very low. However, the models differed when simulating T_a during the plant dominated period from 41 days after planting to 20 days before observed maturity and whole-season ET_a. Considering all 8 years, XNSM, the model with the more detailed root description but a simpler approach in calculating T_p, clearly outperformed the model with more detail in the simulation of atmospheric demand, XNGM. In particular in the typical rainfall year 2011, XNSM simulated mid-season T_a very well. However, XNGM was superior to XNSM in simulating mid-season T_a in the exceptional year 2012. Probably, in this particular dry year the uniform distribution of the roots over the whole rooting zone simulated in XNGM facilitated water uptake from deeper soil layers, where even during the drought period, enough water was available to allow transpiration to continue at high rates, as suggested by observations. The limiting root water uptake capacity introduced in XNSM could be a disadvantage here since it prevents the compensation of water uptake in the upper soil by roots in the subsoil. This indicates that a process such as root water uptake compensation or soil moisture redistribution by hydraulic lift could play an important role and should be considered in a more realistic way in crop growth models. The better performance of XNSM in simulating evapotranspiration over all 8 years is also reflected in a slightly higher accuracy in yield predictions compared to XNGM.

6. Consideration of potential ET_p simulation

We also inter-compared potential ET (ET_p) to determine whether the source of the huge variability among the models in daily ET simulations (Figs. 1 and 2) could be due to the models' approaches for computing ET_p or whether it was more due to later restrictions from plant and/or soil factors going from potential ET_p to actual ET_a. A few models do not use the concept of potential ET_p, but the many modelers who do use the concept reported their ET_p values, and they are plotted in Fig. 10 for Phase 1. From the lower to the higher outliers, there is a huge range, just as in Fig. 1b for actual ET_a. Thus, indeed much of the variability among the models was due to what they computed for ET_p. It appears that large improvements in the poorer models can be

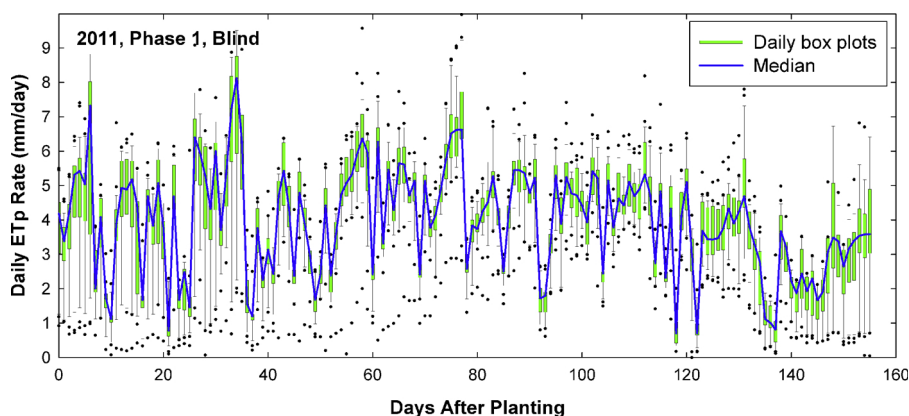


Fig. 10. Box plots of daily potential evapotranspiration (ET_p) reported by the modelers for 2011, Phase 1. Median values from the 29 models are also shown. This graph corresponds to Fig. 1b for actual daily evapotranspiration, ET_a.

achieved by simply adopting the ETp methodology of the better models.

7. Conclusions with discussion

1. There was a huge range among the 29 models in their simulations of daily ET for the initial blind phase and continuing even after the modelers were supplied all available ET and growth observations in subsequent phases (Figs. 1 and 2). Nevertheless, the medians were generally close to observations except during dry periods, and even during these periods, agreement improved with increasing information for calibration. There was relatively little improvement with increasing information during the first 40 days after sowing. In contrast, there was considerable improvement in ET prediction in mid- to late-season in Phase 3 (ET data for 2011 provided) and Phase 4 (all data provided along with allowing water table or increased water-holding capacity at depth). These Phase 3 and 4 improved simulations imply model improvements associated with prediction of either transpiration or of associated soil-root-water-uptake.

2. There were greatly differing approaches among the best seven models for the several categories of inter-comparison, and their accuracies were close to one another and to the median. For example, Aqyield (AQY) is a simple model that uses a crop coefficient and doesn't even compute leaf area or biomass, yet as befitting its name, it performed fairly well at predicting ET, and surprisingly, yield. At the other extreme is the JULES (JUL) model, which has an hourly time step and divided the canopy into ten LAI layers for which an energy balance was computed for each. JUL did the best for predicting ET for all eight growing seasons for blind Phase 1. Intermediate in complexity were versions of the widely used DSSAT-CERES-Maize and APSIM-Maize which did well at predicting both ET and agronomic variables.

While both simple and complex models did well in this inter-comparison, to be fair, one would expect the more complex models to be able to handle a wider range of environmental conditions than were exhibited by the dataset used herein, which was obtained in humid Iowa with little drought stress. For example, in irrigated arid areas, canopy temperatures can easily be 5 °C and more below air temperature (e.g. Idso, 1982), and thermal development times follow canopy rather than air temperatures (e.g., Kimball et al., 2012). Thus, those more complex models that can compute and use canopy temperature more likely can simulate accurate thermal crop development times (e.g., Webber et al., 2016).

3. Although the ensemble median was not the best for several of the categories or phases (Figs. 3, 4 and 7–9), it was the overall best for all four phases when considering both ET and agronomic parameters together (Table 3). That the ensemble median outperforms individual models is consistent with previous inter-comparisons, e.g. Asseng et al. (2015).

4. Choices made by experienced users can be as important as approaches within the models in the accuracy of predictions (Table 4). Within this inter-comparison, there were identical versions of the same APSIM model run by two different users with the result that one did better than the other with reversals going from one phase to another.

5. For the case of the daily and hourly MAZSIM models (MZD and MZH), it made very little difference in performance whether the weather data were supplied as actual hourly observations or whether daily values were supplied and hourly values were simulated (Table 4). Thus, good justification is provided for the use of daily weather data for driving simulation models with a savings in data storage by a factor of 24 compared to hourly data.

6. In the case of SIM-PLACE LINTUL5, the older version (SLNH) based on Hargreaves (1975) provided better estimates of ET during the first 40 days after planting (E dominant) than did the newer version based on FAO56 with dual coefficients and with an energy balance that computes canopy (or surface) temperature (SLFT). However, during the T-dominated portion of the season (41 days after planting to 20 before observed maturity), the opposite was true, and SLFT was better most

years, but not dry 2012. Consistent with doing the better predictions of T, SLFT produced better predictions of leaf area and biomass. However, SLNH produced better estimates of yield suggesting that those aspects in SLFT relating to senescence and harvest index need improvement.

7. The original soil water evaporation scheme in the DSSAT-CSM models, which was Ritchie (1972) two-stage, outperformed the newer default routine, which was Suleiman and Ritchie (2003) that has upward movement of vapor from lower layers “directly” to E only, but not T, at least for this dataset (Table 5).

8. With the DSSAT-CERES-Maize model, the Priestly and Taylor (1972) approach was slightly better than FAO56 (Allen et al., 1998), which in turn was slightly better than the ASCE Dual Kc approach (Allen et al., 2005) at predicting ET (Table 5). The reverse order was expected, so this conclusion might not be true in all environments. However, the differences among these approaches were small, at least relative to the differences due to the soil water evaporation approach (Section 5.3).

9. Having a more realistic detailed root distribution in the case of (XNSM) resulted in more accurate predictions of ET during the T dominated part of the season (41 days after planting to 20 days before observed maturity) compared to having a detailed energy balance with canopy temperature but having a root distribution that was uniform over the whole rooting zone (XNGM; Table 6).

10. It appeared that many of the models could not handle water-logged conditions, such as were experienced in 2010, and many could not handle a water table supplying extra water in drought conditions, such as in 2012. However, these conditions were a relatively small portion of this 8-year study.

11. Much of the wide variability among the models in their ability to simulate ET was due to variability in the approaches used to calculate potential ETp.

12. An objective of this inter-comparison among maize models to predict ET was to identify those approaches that were most successful. However, there was a huge range in complexity among the best six approaches, and many of the other models used approaches that seem similar to the best six (Table S1). Thus, why should there be such a huge range in ET estimates among the 29 models with many models exhibiting large over- and under-estimates from the observations? One likely factor could be the number of users over an extended period of time leading to greater testing of the more successful models, whereas several of the models are relatively new and still undergoing development. Indeed, more than once during this exercise, modelers requested that prior submitted results be discarded and new ones used because they had made a significant change in their model. In a recent survey about model calibration (Seidel et al., 2018), half of the respondents used APSIM or DSSAT models and versions of these two models fared well in this inter-comparison. Another factor leading to a large ET prediction range among models could be the choices made by the users with regard to parameter choice and calibration of the models, as illustrated by the difference in accuracy between two users using the same model within this inter-comparison (Section 5.1). Another reason is that in many cases, the ET aspects aren't as high a priority as the agronomic aspects to model developers and users. In the recent Seidel et al. (2018) survey, 90% of respondents reported calibrating yield, but only 57% for ET.

13. This exercise illustrates the need for good datasets to validate crop growth and water use models. Besides accurate ET measurements and accompanying weather, management, soil property, and final yield data, such datasets also should include shallow and deep soil moisture measurements, initial conditions, and in-season biomass on several dates and final biomass.

Acknowledgements

This experiment would not have been possible without the dedication of Christian Dold, Forrest Goodman, Wolfgang Oesterreich, Laura

Hanse, Michelle Cryder, and several students who assisted in making the measurements and processing the observed data during the time. A portion of this work was conducted by AS and TG under the project BiomassWeb of the GlobeE programme (Grant number: FKZ031A258B) funded by the Federal Ministry of Education and Research (BMBF, Germany). SA contribution was supported by a FFAR grant entitled “Improving simulation of soil water dynamics and crop yields in the corn belt, USA”. MPH acknowledges support from the ‘Limpopo Living Landscapes’ project within the SPACES programme (grant number 01LL1304A) funded by the German Federal Ministry of Education and Research (<http://www.bmbf.de/en/http://www.bmbf.de/en/>).

Any trade, product, or firm name is used for descriptive purposes only and does not imply endorsement by the U.S. Government.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.agrformet.2019.02.037>.

References

- Allen, R.G., Pereira, L.S., Raes, D., Smith, M., 1998. Crop Evapotranspiration: Guidelines for Computing Crop Water Requirements, FAO Irrigation and Drainage Paper 56. Food and Agriculture Organization of the United Nations, Rome, Italy.
- Allen, R.G., Walter, I.A., Elliott, R., Howell, T., Itenfisu, D., Jensen, M., 2005. The ASCE Standardized Reference Evapotranspiration Equation. American Society of Civil Engineers, Reston, Virginia, pp. 195.
- 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., Jones, C.D., Priesack, E., Ripoche, D., Semenov, M.A., 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. *Nat. 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., Jongschaap, R., Kemanian, A.R., Kersebaum, K.C., Koehler, A.K., Müller, 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. *Nat. Clim. Change* 5, 143–147.
- Basso, 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? *Glob. Change Biol.* 20, 2301–2320. <https://doi.org/10.1111/gcb.12520>.
- Boote, K.J., Sau, F., Hoogenboom, G., Jones, J.W., 2008. Experience with water balance, evapotranspiration, and prediction of water stress effects in the CROPGRO model. In: Ahuja, L.R., Reddy, V.R., Saseendran, S.A., Yu, Q. (Eds.), *Response of Crops to Limited Water: Modeling Water Stress Effects on Plant Growth Processes*, Volume 1 of *Advances in Agricultural Systems Modeling*. ASA-CSSA-SSSA, Madison, WI, pp. 59–103.
- Brisson, N., Seguin, B., Bertuzzi, P., 1992. Agrometeorological soil water balance for crop simulation models. *Agric. For. Meteorol.* 59, 267–287.
- 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., Bussièrre, F., Cabidoche, Y.M., Cellier, P., Debaeke, P., Gaudillère, J.P., Hênault, C., Maraux, F., Sequin, B., Sinoquet, H., 2003. An overview of crop model STICS. *Eur. J. 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., Izaurralde, R.C., Müller, C., Kumar, S.N., Nendel, C., O’Leary, G., Olesen, J.E., Osborne, T.M., Priesack, E., Ripoche, D., Steduto, P., Stöckle, C., O., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Travasso, M., Waha, K., White, J.W., Wolf, J., 2016. Uncertainty of wheat water use: simulated patterns and sensitivity to temperature and CO₂. *Field Crops Res.* 198, 80–92. <https://doi.org/10.1016/j.fcr.2016.08.015>.
- Campbell, G.S., 1985. *Soil Physics With BASIC*. Elsevier, New York, New York, pp. 150.
- 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.
- Dold, C., Büyükcangaz, H., Rondinelli, W., Prueger, J.H., Sauer, T.J., Hatfield, J.L., 2016. Long-term carbon uptake of agro-ecosystems in the Midwest. *Agric. For. Meteorol.* 232, 128–140.
- 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 CO₂ concentration levels with limited water supply on water use and yield? *Eur. J. Agron.* 100, 67–75.
- Fan, Y., Li, H., Miguez-Macho, G., 2013. Global patterns of groundwater table depth. *Science* 339, 940–943.
- Fleisher, D.H., Condoni, B., Quiroz, R., Alva, A., Asseng, S., Barreda, C., Bindi, M., Boote, K.J., Ferrise, R., Franke, A.C., Govindkrishnan, 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. *Glob. Change Biol. Bioenergy* 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 Micrometeorology: A Simulation Study*. Simulation Monographs. PUDOC, Wageningen, the Netherlands.
- Goudriaan, J., Van Laar, H.H., 1994. *Radiation in crops*. In: Goudriaan, J., Van Laar, H.H. (Eds.), *Modelling Potential Crop Growth Processes*. Textbook With Exercises. Kluwer Academic Publ., Dordrecht, The Netherlands, pp. 378–399.
- Hamon, W.R., 1963. Computation of direct runoff amounts from storm rainfall. *Int. Assoc. Sci. Hydrol. Publ.* 63, 52–62.
- Hargreaves, G.H., 1975. Moisture availability and crop production. *Trans. ASAE* 18, 980–984.
- 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 CO₂ examined with free-air CO₂ enrichment and growth chamber experiments. *Sci. Rep.* 7, 14858. <https://doi.org/10.1038/s41598-017-13582-y>.
- Hernandez-Ramirez, G., Hatfield, G.L., Prueger, J.H., Sauer, T.J., 2010. Energy balance and turbulent flux partitioning in a corn-soybean rotation in the Midwestern U.S. *Theor. Appl. Climatol.* 100, 79–92.
- Huth, N.I., Bristow, K.L., Verburg, K., 2012. *Swim3*. *Trans. ASABE* 55, 1303–1313.
- Hutson, J.L., Wagenet, R.J., 1992. LEACHM: Leaching Estimation and Chemistry Model: A Process-Based Model of Water and Solute Movement, Transformations, Plant Uptake and Chemical Reactions in the Unsaturated Zone. Version 3.0. Research Series No. 93-3. Cornell University, Ithaca, N. Y.
- Idso, S.B., 1982. Non-water-stressed baselines: a key to measuring and interpreting plant water stress. *Agric. Meteorol.* 27, 59–70.
- Johnsson, H., Bergström, H.L., Jansson, P.E., Paustian, K., 1987. Simulated nitrogen dynamics and losses in a layered agricultural soil. *Agric. Ecosyst. Environ.* 18, 333–356.
- Kim, S.-H., Yang, Y., Timlin, D.J., Fleisher, D., Dathe, A., Reddy, V.R., 2012. Modeling nonlinear temperature responses of leaf growth, development, and biomass in MAIZSIM. *Agron. J.* 104, 1523–1537.
- Kimball, B.A., White, J.W., Wall, G.W., Ottman, M.J., 2012. Infrared-warmed and unwarmed wheat vegetation indices coalesce using canopy-temperature-based growing degree days. *Agron. J.* 104, 114–118.
- 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. *Glob. Change Biol.* 21, 1328–1341.
- Logsdon, S.D., Hernandez-Ramirez, G., Hatfield, J.L., Sauer, T.J., Prueger, J.H., Schilling, K.E., 2009. Soil water and shallow groundwater relations in an agricultural hillslope. *Soil Sci. Soc. Am. J.* 73, 1461–1468.
- Maiorano, A.P., Martre, P., Asseng, S., Ewert, F., Müller, C., Rötter, R.P., Ruane, A.C., Semenov, 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., Priesack, 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.
- Monteith, J.L., 1965. *Evaporation and environment*. 19th Symposia of the Society for Experimental Biology, vol. 19. University Press, Cambridge, pp. 205–234.
- Nimah, M., Hanks, R.J., 1973. Model for estimating soil-water-plant-atmospheric relation: I. Description and sensitivity. *Soil Sci. Soc. Am. Proc.* 37, 522–527.
- Ordóñez, R.A., Castellano, M.J., Hatfield, J.L., Helmers, M.J., Licht, M.A., Liebman, M., Ditzel, R., Martinez-Feria, R., Iqbal, J., Puntel, L.A., Córdova, S.C., Togliatti, K., Wright, E.E., Archontoulis, S.V., 2018. Maize and soybean root front velocity and

- maximum depth in Iowa, USA. *Field Crops Res.* 215, 122–131.
- Pickering, N.B., Jones, J.W., Boote, K.J., 1995. Adapting SOYGRO V5.42 for prediction under climate change conditions. In: Rosenzweig, C., Jones, J.W., Allen Jr.L.H. (Eds.), *Climate Change and Agriculture: Analysis of Potential International Impacts*, ASA Spec. Pub. No. 59. ASA-CSSA-SSSA, Madison, WI, pp. 77–98.
- Priestly, C.H.B., Taylor, R.J., 1972. On the assessment of surface heat flux and evaporation using large-scale parameters. *Mon. 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](https://doi.org/10.1016/S0308-521X(97)00028-0).
- Rippey, B.R., 2015. The U.S. drought of 2012. *Weather Clim. Extreme* 10, 57–64.
- Ritchie, J.T., 1972. Model for predicting evaporation from a row crop with incomplete cover. *Water Resour. Res.* 8, 1204–1213.
- Ritchie, J.T., Godwin, D.C., Otter-Nache, S., 1988. CERES-Wheat: A Simulation Model of Wheat Growth and Development. Texas A&M University Press, College Station, TX.
- Ritchie, J.T., Porter, C.H., Judge, J., Jones, J.W., Suleiman, A.A., 2009. Extension of an existing model for soil water evaporation and redistribution under high water content conditions. *Soil Sci. Soc. Am. J.* 73, 792–801.
- Rizzo, G., Edreira, J.I.R., Archontoulis, S.V., Yang, H.S., Grassini, P., 2018. Do shallow water tables contribute to high and stable maize yields in the US Corn Belt? *Glob. Food Sec.* 18, 27–34.
- Sauer, T.J., 2002. Methods of soil analysis—Part 1, physical and mineralogical methods. In: Dane, J.D., Topp, G.C. (Eds.), *Heat flux density*. American Society of Agronomy, Madison, WI, pp. 1233–1248.
- 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>.
- Shuttleworth, W.J., Wallace, J.S., 1985. Evaporation from sparse crops—an energy combination theory. *Q. J. R. Meteorol. Soc.* 111, 839–855.
- Simunek, 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. Tech. Rep. 144, U.S. Salinity Lab., United States Dep. of Agriculture, Agricultural Research Service.
- Soufzadeh, S., Munaro, E., McLean, G., Massignam, A., van Oosterom, E.J., Chapman, S.C., Messina, C., Cooper, M., Hammer, G.L., 2018. Modeling the nitrogen dynamics of maize crops-enhancing the APSIM maize model. *Eur. J. Agron.* 100, 118–131.
- Suleiman, A.A., Ritchie, J.T., 2003. Modeling soil water redistribution during second-stage evaporation. *Soil Sci. Soc. Am. J.* 67, 377–386.
- Tanner, C.B., Sinclair, T.R., 1983. Efficient water use in crop production: research or research? In: Taylor, H.M., Jordan, W.R., Sinclair, T.R. (Eds.), *Limitations to Efficient Water Use in Crop Production*. American Society of Agronomy, Madison, WI, USA, pp. 1–27.
- Tao, F., Yokozawa, M., Zhang, Z., 2009a. Modelling the impacts of weather and climate variability on crop productivity over a large area: a new process-based model development, optimization, and uncertainties analysis. *Agric. For. Meteorol.* 149, 831–850.
- Timlin, D.J., Pachepsky, Ya., Acocck, B.A., Šimunek, J., Flerchinger, G., Whisler, F., 2002. Error analysis of soil temperature simulations using measured and estimated hourly weather data with 2DSOIL. *Agric. Sys.* 72, 215–239.
- Twine, T.E., Bryant, J.J., Richter, K.T., Bernacchi, C.J., McConnaughay, K.D., Morris, S.J., Leakey, A.D.B., 2013. Impacts of elevated CO₂ concentration on the productivity and surface energy budget of the soybean and maize agroecosystem in the Midwest USA. *Glob. Change Biol.* 19, 2838–2852. <https://doi.org/10.1111/gcb.12270>.
- Villalobos, F.J., Fereres, E., 1990. Evaporation measurements beneath corn, cotton, and sunflower canopies. *Agron. J.* 82, 1152–1159 3-2.
- 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>.
- Webb, E.K., Pearman, G.I., Leuning, R., 1980. Correction of flux measurements for density effects due to heat and water vapour transfer. *Q.J.R. Meteorol. Soc.* 106, 85–100.
- Webber, H., Ewert, F., Kimball, B.A., Siebert, S., White, J.W., Wall, G.W., Ottman, M.J., Trawally, D.N.A., Gaiser, T., 2016. Simulating canopy temperature for modelling heat stress in cereals. *Environ. Model. Softw.* 77, 143–155. <https://doi.org/10.1016/j.envsoft.2015.12.003>.
- 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.
- Yin, X., van Laar, H.H., 2005. *Crop Systems Dynamics: An Ecophysiological Simulation Model for Genotype-By-Environment Interactions*. Wageningen Academic Publishers.