Simulation of evapotranspiration and yield of maize

Bruce Kimball, Kelly Thorp, Kenneth Boote, Claudio Stockle, Andrew Suyker, Steven Evett, David Brauer, Gwen Coyle, Karen Copeland, Gary Marek, et al.

HAL Id: hal-04112537
https://hal.inrae.fr/hal-04112537v1

Submitted on 28 Aug 2023

Simulation of Evapotranspiration and Yield of Maize:

An Inter-comparison among 41 Maize Models

By

Bruce A. Kimball[1], Kelly R. Thorp[1], Kenneth J. Boote[2], Claudio Stockle[3], Andrew E. Suyker[4], Steven R. Evett[5], David K. Brauer[5], Gwen G. Coyle[5], Karen S. Copeland[5], Gary W. Marek[5], Paul D. Colaizzi[5], Marco Acutis[6] Seyyedmajid Alimagham[7], Sotirios Archontoulis[8], Faye Babacar[9], Zoltán Barcza[10,11], Bruno Basso[12], Patrick Bertuzzi[13], Julie Constantin[14], Massimiliano De Antoni Migliorati[15], Benjamin Dumont[16], Jean-Louis Durand[17], Nándor Fodor[11,18], Thomas Gaiser[19], Pasquale Garofalo[20], Sebastian Gayler[21], Luisa Giglio[20], Robert Grant[22], Kaiyu Guan[23], Gerrit Hoogenboom[2], Qianjing Jiang[24], Soo-Hyung Kim[25], Isaya Kisekka[26], Jon Lizaso[27], Sara Masia[28], Huimin Meng[29], Valentina Mereu[30], Ahmed Mukhtar[31], Alessia Perego[6], Bin Peng[23], Eckart Priesack[32], Zhiming Qi[24], Vakhtang Shelia[2], Richard Snyder[33], Afshin Soltani[7], Donatella Spano[30], Amit Srivastava[19], Aimee Thomson[34], Dennis Timlin[35], Antonio Trabucco[30], Heidi Webber[36], Tobias Weber[21], Magali Willaume[14], Karina Williams[37,38], Michael van der Laan[34], Domenico Ventrella[20], Michelle Viswanathan[21], Xu Xu[29], Wang Zhou[23]

[1]U.S. Arid-Land Agricultural Research Center, USDA-ARS, Maricopa, AZ 85138

(bruce.kimball@usda.gov; kelly.thorp@usda.gov)

[2]University of Florida, Agricultural and Biological Engineering, Frazier Rogers Hall,

Gainesville, Florida 32611-0570, USA (kjboote@ufl.edu; gerrit@ufl.edu;

vakhtang.shelia@ufl.edu)

[3]Biological Systems Engineering, Washington State University, 1935 E. Grimes Way, PO Box

646120, Washington State University, Pullman WA 99164-6120 (stockle@wsu.edu)

[4]School of Natural Resources, University of Nebraska-Lincoln, Lincoln, Nebraska, USA

(asuyker1@unl.edu)

[5]Conservation and Production Research Laboratory, USDA-ARS, Bushland, Texas, USA

(Steve.Evett@usda.gov; david.brauer@usda.gov; gwen.coyle@usda.gov;

karen.copeland@usda.gov; gary.marek@usda.gov; paul.colaizzi@usda.gov)

[6]Department of Agricultural and Environmental Sciences, University of Milan, via Celoria 2 –

20133, Milan, Italy (marco.acutis@unimi.it, alessia.perego@unimi.it)

32

[7]Agronomy Group, Gorgan Univ. of Agric. Sci. & Natur. Resour., Gorgan 49138-15739 Iran

(Afshin.Soltani@gmail.com; m.alimagham@gmail.com)

[8]Iowa State University, Department of Agronomy, Ames, Iowa 50010 (sarchont@iastate.edu; 1-

515-294-7413)

[9]Institut de recherche pour le développement (IRD) ESPACE-DEV, F-34093 Montpellier Cedex,

France (babacar.faye@ird.fr)

[10]ELTE Eötvös Loránd University, Department of Meteorology, H-1192 Budapest, Hungary
(zoltan.barcza@ttk.elte.hu)

[11]Czech University of Life Sciences Prague, Faculty of Forestry and Wood Sciences, 165 21

Prague, Czech Republic

[12]Michigan State University, Dept. Geological Sciences and W.K. Kellogg Biological Station,

288 Farm Ln, 307 Natural Science Bldg., East Lansing, MI, 48824 (basso@msu.edu)

[13]US1116 AgroClim, INRAE centre de recherche Provence-Alpes-Côte d'Azur, 228, route de

l'Aérodrome, CS 40 509, Domaine Saint Paul, Site Agroparc, 84914 Avignon Cedex 9, France

(agroclim-contact@inrae.fr)

[14]AGIR, Université de Toulouse, INRAE, INPT, INP- EI PURPAN, 24 Chemin de Borde Rouge

- Auzeville CS 52627, CastanetTolosan, France (julie.constantin@toulouse.inra.fr;

magali.willaume@ensat.fr).

[15]Queensland Department of Environment & Science, Queensland, Australia

(Max.DeAntoni@des.qld.gov.au)

54 [16]ULiege-GxABT, University of Liege - Gembloux Agro-Bio Tech, TERRA Teaching and

55 research centre, Plant Science Axis / Crop Science Lab., B-5030 Gembloux, Belgium

56 (Benjamin.Dumont@uliege.be)

57 [17]Unité de Recherches Pluridisciplinaire Prairies et Plantes Fourragères, INRAE, 86 600

58 Lusignan, France (jean-louis.durand@inra.fr)

59 [18]Agricultural Institute, Centre for Agricultural Research, H-2462 Martonvásár, Brunszvik u. 2.,

60 Hungary (fodor.nandor@atk.hu)

61 [19]Institute of Crop Science and Resource Conservation, University of Bonn, Katzenburgweg 5

62 D-53115 Bonn, Germany (tgaiser@uni-bonn.de; amit.srivastava@uni-bonn.de)

63 [20]Council for Agricultural Research and Economics, Agriculture and Environment Research

64 Center, CREA-AA, Via Celso Ulpiani 5, 70125 BARI BA, Italy (pasquale.garofalo@crea.gov.it;

65 luisa.giglio@crea.gov.it; domenico.ventrella@crea.gov.it)

66 [21]Universitat Hohenheim, Institute of Soil Science and Land Evaluation, Biogeophysics, Emil-

67 Wolff-Str. 27, D-70593 Stuttgart, Germany (sebastian.gayler@uni-hohenheim.de;

68 tobias.weber@uni-hohenheim.de; michelle.viswanathan@uni-hohenheim.de)

69 [22]Department of Renewable Resources, University of Alberta, Edmonton, Alberta, Canada T6G

70 2E3 (rgrant@ualberta.ca)

71 [23]College of Agricultural, Consumer and Environmental Sciences (ACES), University of Illinois

72 at Urbana-Champaign, Urbana, Illinois 61801, USA (kaiyug@illinois.edu;

73 binpeng@illinois.edu; wangzhou@illinois.edu)

74 [24]Department of Bioresource Engineering, Macdonald Campus, McGill University, 1-024

75 Macdonald-Steward Hall, Sanite-Anne-de-Bellevue, QC, Canada H9X 3V9

76 (qianjing.jiang@mail.mcgill.ca; zhiming.qi@mcgill.ca)

[25]School of Environmental and Forest Sciences, University of Washington, Seattle, WA 98195

(soohkim@uw.edu)

[26]Agricultural Water Management and Irrigation Engineering; University of California Davis;

Departments of Land, Air, and Water Resources and of Biological and Agricultural Engineering;

One Shields Avenue; PES 1110; Davis, CA 95616-5270, USA (ikisekka@ucdavis.edu)

[27]Technical University of Madrid (UPM), Dept. Producción Agraria-CEIGRAM, Ciudad

Universitaria, 28040 Madrid, Spain (jon.lizaso@upm.es)

[28]Land and Water Management Department, IHE Delft Institute for Water Education, Delft, The

Netherlands (sara.masia@cmcc.it)

[29]Center of Agricultural Water Research, China Agricultural University, Beijing, China

(mhm2019@163.com; xushengwu@cau.edu.cn)

[30]Fondazione CMCC - Euro-Mediterranean Centre on Climate Change, Impacts on Agriculture,

Forests and Ecosystem Services division (IAFES), Sassari, ITALY (antonio.trabucco@cmcc.it;

spano@uniss.it; valentina.mereu@cmcc.it)

[31]Department of Agronomy, PMAS Arid Agriculture University, Rawalpindi, Pakistan and

Swedish University of Agricultural Sciences, Umea Sweden (mukhtar.ahmed@slu.se)

[32]Helmholtz Center Munich, Institute of Biochemical Plant Pathology, Ingolstaedter Landstr. 1

85764 Neuherberg, Germany (priesack@helmholtz-muenchen.de)

[33]University of California Davis (rlsnyder@ucdavis.edu)

[34]University of Pretoria, Pretoria, South Africa (michael.vanderlaan@up.ac.za;

u16015925@tuks.co.za)

[35]Crop Systems and Global Change Research Unit, USDA-ARS, Beltsville, MD

(Dennis.Timlin@ars.usda.gov)

100   [36]Leibniz Centre for Agricultural Landscape Research (ZALF), Mucheberg 15374, Germany

101   (webber@zalf.de)

102   [37]Hadley Centre, FitzRoy, Road Exeter Devon EX1 3PB, United Kingdom

103   (karina.williams@metoffice.gov.uk)

104   [38]Global Systems Institute, University of Exeter, North Park Road, Exeter, EX4 4QE, UK
105

106

107   Corresponding Author:

108   Bruce A. Kimball

109   U.S. Arid-Land Agricultural Research Center

110   USDA, Agricultural Research Service

111   21881 North Cardon Lane

112   Maricopa, Arizona 85018, USA

113   (phone: 1-602-840-4352; email: bruce.kimball@usda.gov)

114

115   **Abstract**

116   Accurate simulation of crop water use (evapotranspiration, ET) can help crop growth models to

117   assess the likely effects of climate change on future crop productivity, as well as being an aid for

118   irrigation scheduling for today's growers. To determine how well maize (*Zea mays* L.) growth

119   models can simulate ET, an initial inter-comparison study was conducted in 2019 under the

120   umbrella of AgMIP (Agricultural Model Inter-Comparison and Improvement Project). Herein,

121   we present results of a second inter-comparison study of 41 maize models that was conducted

122   using more comprehensive datasets from two additional sites - Mead, Nebraska, USA and

123   Bushland, Texas, USA. There were 20 treatment-years with varying irrigation levels over

124   multiple seasons at both sites. ET was measured using eddy covariance at Mead and using large

125   weighing lysimeters at Bushland. A wide range in ET rates was simulated among the models, yet

126   several generally were able to simulate ET rates adequately. The ensemble median values were

127   generally close to the observations, but a few of the models sometimes performed better than the

128   median. Many of the models that did well at simulating ET for the Mead site did poorly for drier,

129   windy days at the Bushland site, suggesting they need to improve how they handle humidity and

130   wind. Additional variability came from the approaches used to simulate soil water evaporation.

131   Fortunately, several models were identified that did well at simulating soil water evaporation,

132   canopy transpiration, biomass accumulation, and grain yield. These models were older and have

133   been widely used, which suggests that a larger number of users have tested these models over a

134   wider range of conditions leading to their improvement. These revelations of the better

135   approaches are leading to model improvements and more accurate simulations of ET.

136

137   **Graphical abstract**

138
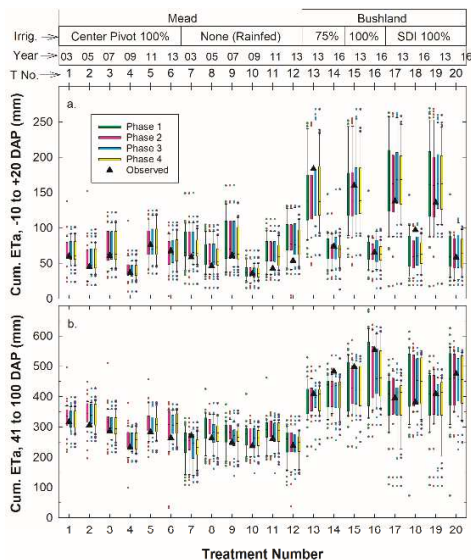
Highlights

- Maize growth models differ widely in their simulations of daily evapotranspiration

- Most models fail to sufficiently reduce transpiration after crop maturation

- Most models fail to adequately reproduce effects of low humidity and high windspeed

- The median of models was often but not always the best performing

- Model inter-comparisons suggest avenues to improve simulation of maize ET

Keywords: Maize; simulation; evapotranspiration; water use; crop models; yield

1 Introduction

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

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

166   et al., 2015; Hasegawa et al., 2017), and potato (*Solanum tuberosum* L.; Fleisher et al., 2017)].

167   As discussed by Kimball et al. (2019), only a few comparisons among methods or models to

168   simulate ET have been done previously. Sau et al. (2004) evaluated several ET options with the

169   CROPGRO Faba bean (*Vicia faba* L.) model, by careful comparison to soil water balance, and

170   found that the FAO-56 option (Allen et al., 1998) had a root mean square error (RMSE) that was

171   20% smaller than the Priestley-Taylor option (Priestley and Taylor, 1972) and 48% smaller than

172   the FAO-24 option (Doorenbos and Pruitt, 1985). In an inter-comparison of water use among 16

173   wheat models at four sites around the world, Cammarano et al. (2016) found the coefficient of

174   variation was only 22.5% among models and sites. In contrast, in an inter-comparison among 23

175   maize models, Bassu et al. (2014) found a very large range of simulated values of ET among the

176   models, including -10 to +30% variations in the ET response to doubled CO_2 concentration (720

177   μmol/mol). However, there were no observations of ET or water use in the dataset chosen for

178   that study, so there was no standard for comparison. Therefore, Kimball et al. (2019) conducted

179   their study using eight seasons of data from Ames, Iowa, USA for which eddy covariance

180   measurements of ET were available. Like Bassu et al. (2014), they also found simulated ET

181   values varied by a factor of two among the maize models. Surprisingly, among the models with

182   closest agreement to observations, some were quite simple (e.g., no simulation of biomass) and

183   some were quite complex (e.g., full energy balance), so it was difficult to determine which

184   approaches were generally best and should be adopted by the poorer performing models.

185   Nevertheless, there were several cases in which different ET methods were tested within the

186   same family of crop models, and comparisons among these methods clearly revealed some

187   approaches that were better than others.

188

189 However, there were some issues with the Ames dataset (Kimball et al., 2019). For example, in

190 2012, an infamous year for drought in the Midwest, observed ET and crop yield were higher than

191 in other years. Further analysis led to the strong suspicion that there was a water table present to

192 provide additional water besides the sparse rainfall, yet deep soil water measurements were

193 lacking to confirm the suspicion. Therefore, it was decided to repeat the study of Kimball et al.

194 (2019) with more robust datasets.

195

196 Two such datasets were identified, one from the University of Nebraska at Mead, Nebraska,

197 USA (41.165°N, 96.470°W, 362 m), which is close to the 100th meridian typically used to divide

198 the humid East from the arid West, thus placing it within the U.S. "corn belt." There were six

199 seasons of maize from irrigated and rainfed fields (12 treatment-years) with ET determined using

200 eddy covariance. The second was collected by the USDA, Agricultural Research Service,

201 Conservation and Production Research Laboratory (CPRL), Bushland, Texas, USA (35.183°N,

202 102.100°W, 1170 m), which is a more arid region where maize is mostly grown with irrigation,

203 and where winds are commonly higher. They measured ET using large weighing lysimeters.

204 They grew maize for two seasons with MESA (mid-elevation sprinkler application) at 100% and

205 75% replacement of soil water and in near-duplicate SDI (sub-surface drip irritation) fields at

206 100% (8 treatment-years). A total of 41 models participated in this second round of maize ET

207 simulation inter-comparisons (Tables 1, S1), and again the primary objective was to identify the

208 approaches that were most accurate for simulating ET, i.e., had the lowest RMSE compared to

209 the observations. Besides ET, other objectives were to test the models' abilities to simulate LAI,

210 biomass, grain yield, soil moisture, and soil temperature. By "approaches" we mean the methods

9

211   used by the models to simulate ET or other processes, i.e., FAO-56 (Allen et al. 1998) versus

212   Priestley-Taylor (1972), etc.

213

214

215   **2 Materials and Methods**

216   2.1 *Observed data*

217   2.1.1 *University of Nebraska, Mead, Nebraska, USA*

218   One set of field data came from the University of Nebraska Agricultural Research and

219   Development Center near Mead, Nebraska, USA (http://csp.unl.edu/public/). The soils were deep

220   silty clay loams of Yutan (fine-silty, mixed, superactive, mesic Mollic Hapludalfs), Tomek (fine,

221   smectitic, mesic Pachic Argialbolls), Filbert (fine, smectitic, mesic Vertic Argialbolls), and

222   Filmore (fine, smectitic, mesic Vertic Argialbolls). The eddy covariance technique was used to

223   determine ET of maize and soybean (*Glycine max*) in alternate years, as well as fluxes of

224   sensible heat and CO_2. Additional details can be found in Suyker and Verma (2008, 2009) and

225   Suyker et al. (2004, 2005). Briefly, fluxes of latent heat, sensible heat, and momentum were

226   determined using data from the following sensors at each site: an omnidirectional 3D sonic

227   anemometer (Model R3: Gill Instruments Ltd., Lymington, UK) and an open-path infrared

228   CO_2/H_2O gas analyzing system (Model LI7500: Li-Cor Inc., Lincoln, NE).

229

230   The instruments were deployed near the centers of the fields, and the fetch was about 400 m in

231   all directions. The eddy covariance sensors were mounted 5.5 m above the ground. Fluxes were

232   corrected for inadequate sensor frequency, and they were also adjusted for the variation in air

233   density due to the transfer of water vapor and sensible heat. Air temperature and relative

234   humidity (Humitter50Y, Vaisala, Helsinki, FIN), soil heat flux at 0.06m (Radiation and Energy

235   Balance Systems, Inc., Seattle, WA), and net radiation at 5.5m (CNR1, Kipp and Zonen Ltd.,

236   Delft, NLD) were also measured. Missing data due to sensor malfunction, power outages,

237   unfavorable weather, etc. (approximately 15–20% per year), were estimated using an approach

238   that combined measurement, interpolation, and empirical data synthesis. When hourly values

239   were missing (day or night), the latent heat values were estimated as a function of available

240   energy. Linear regressions between latent heat and available energy were determined (separately

241   for dry and wet conditions) for sliding 3-day intervals, and these estimates were used to fill in

242   missing flux values.

243

244   To check closure of the energy balance, the sum of latent and sensible heat fluxes ($\lambda E + H$)

245   measured by eddy covariance were plotted against the sum of R_n (net radiation) + four storage

246   terms, determined by other methods (e.g., Suyker and Verma, 2008). Linear regressions were

247   calculated between the hourly values of $H + \lambda E$ and $R_n + G$ at the study sites (excluding winter

248   months and periods with rain and irrigation). Here $G = G_s$ (soil heat storage) + G_c (canopy heat

249   storage) + G_m (heat stored in the mulch) + G_p (energy used in photosynthesis). The regression

250   slopes averaged 0.89 ± 0.08, implying a fairly good closure of the energy balance.

251

252   We used values of daily ET flux, called observed-ET for 2003, 2005, 2007, 2009, 2011, and

253   2013 from the US-Ne2 (41.165° N, 96.470° W, 362 m; http://ameriflux.lbl.gov/sites/siteinfo/US-

254   Ne2) irrigated maize-soybean rotation field and from the US-Ne3 (41.180° N, 96.440° W, 363

255   m; http://ameriflux.lbl.gov/sites/siteinfo/US-Ne3) rainfed maize-soybean rotation field.

256   Conservation tillage practices were used, so plant residues were not ploughed into the soil, and

257   the soil surface was generally partially covered with prior soybean crop residue. Both sites are

258   part of the Ameriflux (https://ameriflux.lbl.gov/sites) U.S. surface gas flux observation system,

259   and the two sites are within 1.6 km of each other. The cultivars were Pn33B51, Pn33G66,

260   Pn33H26, Pn33T57, DK_61-72, and DK_62-98 used in 2003, 2005, 2007, 2009, 2011, and 2013,

261   respectively. The irrigated crops were planted on 14 May, 2 May, 1 May, 21 April, 17 May, and

262   30 May, and the rainfed crops on 13 May, 26 April, 2 May, 22 April, 2 May, and 13 May in

263   2003, 2005, 2007, 2009, 2011, and 2013, respectively. Destructive measurements of green leaf

264   area index (LAI) and biomass were made approximately bi-monthly during the growing season.

265

266   *2.1.2 USDA, Agricultural Research Service, Conservation and Production Research Laboratory,*

267   *Bushland, Texas, USA*

268   Maize was grown in 2013 and 2016 at the USDA-ARS Conservation and Production Research

269   Laboratory (https://www.ars.usda.gov/plains-area/bushland-tx/cprl/), Bushland, Texas (35.18° N,

270   102.10° W, 1170 m above MSL) on a gently sloping (<0.3%) Pullman soil (fine, mixed,

271   superactive, thermic Torrertic Paleustoll). Additional details and data are provided by Evett et al.

272   (2019, 2020, 2022). Four 4.4 ha fields, approximately square in shape and adjacent to each other,

273   each contained a large (3 m × 3 m in surface area, 2.3 m deep) precision weighing lysimeter in

274   the center. The lysimeters contained undisturbed cores of the Pullman soil obtained on site, and

275   they had an accuracy of 0.04 mm water depth equivalent or better (Evett et al., 2012; Marek et

276   al., 1988). The fields and their associated lysimeters were designated NE, SE, NW, and SW

277   according to the inter-cardinal directions. The NE and SE lysimeters and fields were irrigated by

278   subsurface drip irrigation (SDI), and the NW and SW lysimeters and fields were irrigated by

279   mid-elevation sprinkler application (MESA) using a ten-span linear-move system described by

280   Evett et al. (2019). Adaptation of SDI for the NE and SE weighing lysimeters was described by

281   Evett et al. (2018a). A 109-day drought-tolerant variety (Pioneer 1151AM AquaMax, ≤80% Bt)

282   was planted on 16-17 May 2013 under MESA irrigation, on 22-23 May 2013 in the SDI fields

283   and on 10-11 May 2016 in all fields. These are typical dates for maize planting in the region.

284   Crops were managed and fertilized for high grain yield, as detailed by Evett et al. (2019). In each

285   field, destructive subsampling for leaf area index and biomass occurred in replicate plots

286   periodically during the season, and plant height and row width were measured at the same times.

287   Maize harvests were on 15 October 2013 and on 13 and 17 October 2016.

288

289   Soil water content was sensed at center depths of 0.10 to 2.30 m in 0.20 m increments in each of

290   eight access tubes in the field around each lysimeter and in two access tubes in each lysimeter (to

291   1.90 m depth) on a weekly basis, unless prevented by wet field conditions, using a field-

292   calibrated neutron probe and depth-control stand (Evett et al., 2008). Once the crop was

293   established, irrigations were scheduled weekly to replenish the soil water in the top 1.5 m of the

294   profile to field capacity (i.e., replenishing 100% of crop ET), except for one MESA field where

295   irrigations were 75% of full crop ET after crop establishment. As explained by Evett et al.

296   (2019), the MESA 75% deficit irrigation treatment was established to complete a previous

297   longer-term deficit irrigation study. In some cases, two or even three irrigations were required in

298   a week to replenish the water used by the crop. Irrigations by sprinkler and by SDI typically did

299   not occur on the same day. Neutron probe readings were delayed until the soil surface was dry

300   enough to walk on. The soil profile in early 2013 was quite dry, and SDI preplant irrigation and

301   SDI irrigation immediately after planting were required to plant and germinate the crop. This

302   resulted in a full soil profile in the SDI fields by the time neutron probe sensing began, while

303   crop germination with MESA irrigation was accomplished with less frequent irrigations that did

304   not penetrate to the 1.5 m depth. Irrigations in the 100% SDI and MESA fields maintained the

305   soil water depletion to less than the management-allowed depletion level throughout the season.

306   In 2016, the soil profile was much wetter following a wet winter, and no preplant irrigation and

307   less irrigation immediately after planting were needed. Again, irrigations in the 100% SDI and

308   MESA fields kept soil water depletion to less than the management-allowed depletion level.

309

310   Evapotranspiration (ET) was determined on 5 min, 15 min, and daily bases using data analyses

311   and quality control procedures described by Marek et al. (2014) and Evett et al. (2019). Fifteen-

312   minute-average weather data were output from the research weather station of the USDA-ARS

313   Soil and Water Management Research Unit at Bushland, Texas located immediately east of the

314   lysimeter fields. The weather station instrumentation and data quality assurance and control

315   procedures were applied as described by Evett et al. (2018b).

316
317   *2.2 Modeling methodology*

318   2.2.1 *Model list.* The simulations were conducted by 20 modeling groups from around the world

319   with 41 models completing the inter-comparison (Table 1). Details about each model are

320   presented in supplementary Table S1. However, as can be seen from the names (Tables 1, S1), in

321   some cases there were several "flavors" of different simulation methods tested within the same

322   model family that were chosen by the user at run time. The biggest example is that of the

323   DSSAT family (Hoogenboom et al., 2019a,b; Jones et al., 2003) of the Cropping System Model

324   (CSM) within which both the CSM-CERES-Maize and CSM-IXIM-Maize (hereafter simply

325   called CERES and IXIM) modules were run. Both calculate a value called potential

326   evapotranspiration, ETp, which was done using four methods: (1) FAO-56 (Allen et al., 1998),

14

327   (2) Priestley-Taylor (1972), (3) the ASCE Standardized Reference Evapotranspiration Equation

328   (Allen et al., 2005) for 12-cm grass (short crop), and (4) the ASCE Equation for 50-cm alfalfa

329   (tall crop; *Medicago sativa* L.) with FAO-56 dual crop coefficients for maize (Table S1). Within

330   these eight combinations, two E methods for calculating soil water evaporation were tested:

331   "Ritchie" (Ritchie, 1972) and "Suleiman" (Suleiman and Ritchie, 2003, 2004). In addition,

332   within the CERES-FAO-56 and CERES-Priestley-Taylor combinations, E was also computed

333   using Hydrus (Šimůnek et al., 1998, 2008; Shelia et al., 2018), in which soil water moves based

334   on potential gradients. Thus, there were a total of 18 (2 models x 4 ETp methods x 2 soil E

335   methods + 2 Hydrus) DSSAT flavors. Within the DSSAT flavors, model calibrations though

336   Phase 4 were aimed at the best statistics [lowest RMSE, and highest D-statistic (Willmont,

337   1982)] for growth, grain yield, ET, and soil water variables, averaged over four ET options (two

338   ET by two E methods) in order to minimize bias. The ASCE and Hydrus ET options were not

339   included in this process because the methods were not part of the DSSAT V4.7 release, so they

340   were at a slight disadvantage because they were not independently calibrated. Nevertheless, the

341   resulting cultivar coefficients were consistently used among all the DSSAT simulations.

342
343   In addition, Expert-N had GECROS and SPASS flavors, STICS had KETP and ETP_SW

344   flavors, and MAIZSIM had daily and hourly flavors.

345

346

347 Table 1. List of models and their acronyms. (For details about the evapotranspiration aspects of
348 each, see Supplementary Table S1: List of Models Plus Their Simulation Characteristics and
349 Comparisons of Soil Moisture Simulations)
350

Acronym	Model Name	Reference
AHC	Agro-Hydrological & chemical & Crop sys. simulator	Williams et al., 1989
AMSW	APSIM-SOILWAT	Keating et al., 2003
AQCP	AquaCrop	Allen et al., 1998
AQY	Aqyield	Constantin et al., 2015
ARMO	ARMOSA	Perego et al., 2013
BIOM	Biome-BGCMuSo 6.0.2	Hidy et al., 2016
CS	CropSyst4	Stöckle et al., 2003
DACT	DayCent-CABBI	Moore et al., 2020
DCAR	DSSAT CSM-CERES-Maize ASCE-Alfalfa Ritchie	DeJonge & Thorp, 2017
DCAS	DSSAT CSM-CERES-Maize ASCE-Alfalfa Suleiman	DeJonge & Thorp, 2017
DCFH	DSSAT CSM-CERES-Maize FAO-56 Hydrus	Shelia et al., 2018
DCFR	DSSAT CSM-CERES Maize FAO-56 Ritchie	Sau et al., 2004
DCFS	DSSAT CSM-CERES-Maize FAO-56 Suleiman	Sau et al., 2004
DCGR	DSSAT CSM-CERES-Maize ASCE-Grass Ritchie	DeJonge & Thorp, 2017
DCGS	DSSAT CSM-CERES-Maize ASCE-Grass Suleiman	DeJonge & Thorp, 2017
DCPH	DSSAT CSM-CERES-Maize Priestley-Taylor Hydrus	Shelia et al., 2018
DCPR	DSSAT CSM-CERES-Maize Priestley-Taylor Ritchie	Sau et al., 2004
DCPS	DSSAT CSM-CERES-Maize Priestley-Taylor Suleiman	Sau et al., 2004
DIAR	DSSAT CSM-IXIM-Maize ASCE-Alfalfa Ritchie	DeJonge & Thorp, 2017
DIAS	DSSAT CSM-IXIM-Maize ASCE-Alfalfa Suleiman	DeJonge & Thorp, 2017
DIFR	DSSAT CSM-IXIM-Maize FAO-56 Ritchie	Sau et al., 2004
DIFS	DSSAT CSM-IXIM-Maize FAO-56 Suleiman	Sau et al., 2004
DIGR	DSSAT CSM-IXIM-Maize ASCE-Grass Ritchie	DeJonge & Thorp, 2017

16

Acronym	Model Name	Reference
DIGS	DSSAT CSM-IXIM-Maize ASCE-Grass Suleiman	DeJonge & Thorp, 2017
DIPR	DSSAT CSM-IXIM-Maize Priestley-Taylor Ritchie	Sau et al., 2004
DIPS	DSSAT CSM-IXIM-Maize Priestley-Taylor Suleiman	Sau et al., 2004
ECOS	ecosys	Grant & Flanagan, 2007
JUL	JULES	Best et al., 2011
L5SH	L5-SLIM-H	Wolf, 2012
MZD	MAIZSIM Daily	Yang et al., 2009
MZH	MAIZSIM Hourly	Yang et al., 2009
SLUS	SALUS	Basso & Ritchie, 2015
SLFT	SIMPLACE LINTUL5 FAO56 SLIM3 CanopyT	Wolf, 2012
SMET	SIMETAW#	Mancosu et al 2016
SSMi	SSM-iCROP	Soltani & Sinclair, 2012
STCK	STICS_KETP	Brisson et al., 2003
STSW	STICS_ETP_ SW	Brisson et al., 2003
SWB	SWB	Annandale et al., 2000
TMOD	Test Model	
XNGM	Expert-N - GECROS	Priesack et al., 2006
XNSM	Expert-N - SPASS	Priesack et al., 2006

351

352 2.2.2 *Simulation protocol.* The study was conducted in four phases:

1. "Blind phase." The modelers were sent key input data about soils, weather, and management (planting dates, irrigations, fertilizer applications, etc.) information. They also received anthesis and maturity dates, but no other information about plant growth, grain yield, or water use.

2. "Potential or non-stressed growth phase". The modelers were sent time-series leaf area index (LAI) and biomass observations, as well as final grain yields for all the non-water-stress treatments (only irrigated for Mead; only 100% irrigation for Bushland)

3. "Non-stress ET phase". The modelers were sent all ET, soil water, and soil temperature for the non-water-stress treatments (only irrigated for Mead; only 100% irrigation for Bushland)

4. "All phase". In this final phase, the modelers were provided with all LAI, biomass, grain yield, ET, soil moisture, soil temperature etc. data for all treatment-years.

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

371 2.2.3 *Methods for evaluating model performance*

372 Correlation coefficients (r), D statistics (Willmott, 1982), root mean squared errors between observed and simulated values (RMSE), normalized root mean squared errors (nRMSE), average differences, as well as mean squared deviations (MSD), standard bias (SB), non-unity of slopes

18

375   (NU), and lack of correlations (LC) following Gauch et al. (2003), are all presented as

376   Supplementary Statistical Data for Phase 4. Also included are slopes and intercepts of

377   regressions of observed on simulated data, along with corresponding graphs for each model and

378   analyzed parameter.

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

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

381   where n = number of observations, P_i and O_i are the simulated and observed ith value pair, and $\bar{O}$

382   is the observed mean. Normalizing with $\bar{O}$ enables a comparison of the variability of parameters

383   with widely different units and scales, such as ET rate and biomass accumulation, although

384   admittedly, nRMSE fails for $\bar{O}$ = zero or small values close to zero.

385   2.2.4 *k-means clustering*

386   A *k*-means clustering algorithm was used to group models with similar nRMSE statistics and

387   identify the top-performing models in a non-arbitrary way. Analyses focused on nRMSE for four

388   pairs of model output variables, including simulated ET (ETs) versus grain yield and biomass

389   from -10 to +20 days after planting (DAP) (soil E dominant) and from 41 to 100 DAP (canopy T

390   dominant). Initial tests varied the number of clusters (*n*) from 1 to 19. The final analysis was

391   conducted with *n*=4 clusters based on reducing the sum of squared distance from the cluster

392   center to less than 20% of that for *n*=1 cluster. Using *n*=4 clusters also resulted visually

393   appealing cluster plots with the set of top-performing models clearly identified within groups

394   having low nRMSE for both variables of each pair. The *k*-means analysis was conducted using

395   the "scikit-learn" package for Python (Pedregosa et al., 2011).

396

397

3 Simulation results and discussion

3.1 *Daily results for irrigated and rainfed Mead in 2003 (the driest year) and Bushland 100%*

and 75% sprinkler irrigations in 2013 (the year with highest observed ET rates)

These four cases were selected from among the twenty treatment-years available for more

detailed (daily) examination because they represent the two sites and the two cases at each site

with the likely greatest water stress difference between treatments, i.e., irrigated versus rainfed in

the driest year at Mead and 100% versus 75% MESA irrigation in the year with the highest daily

ET rates in Bushland.

3.1.1 *Daily simulated evapotranspiration (ETs)*

3.1.1.1 *Irrigated Mead in 2003.* As found previously (Kimball et al., 2019), there was a wide

range in ETs among the models (Fig. 1). However, the median of all the models tended to be

close to the observed values most days. Admittedly, for this intercomparison, as well as for all

the others that follow in the rest of this paper, the median is biased toward DSSAT because of

the large number of "flavors." For this case, the observed values fell within the short (1-3 mm/d)

green boxes much of the time, which indicates many of the models produced respectable

simulations. There was only a slight (< 1 mm/d) improvement in model performance going from

Phase 1 to Phase 4. It appears that the greatest variability and uncertainty among the models

occurred from about 10 days before planting to 10 DAP (soil E dominated) and from about 130

to 160 DAP (after the crop matured). A likely cause of the latter issue is that many models

retained a fair amount of green LAI at and after simulated maturity; thus, model equations for ET

that depend strongly on LAI did not result in sufficient termination of ET. Successive model

adjustments or calibrations going from Phase 1 to Phase 4 as more information was provided

only slightly improved this. Model code improvement is needed to decrease green LAI due to

20

421   senescence, eventually shutting down T at crop maturity. Code improvement likely is also

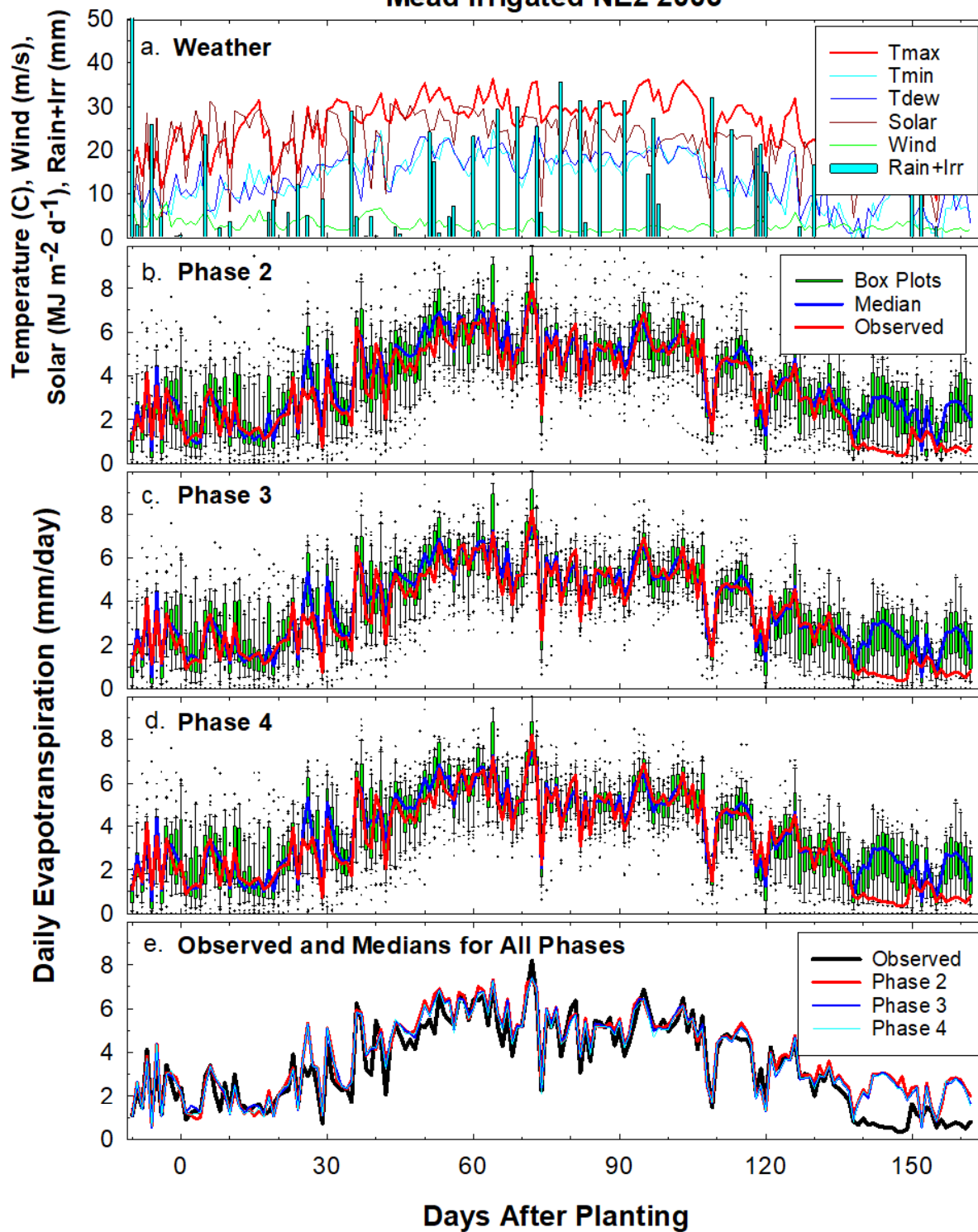422   needed to improve the simulation of bare soil ET.

423   3.1.1.2 *Rainfed Mead in 2003.* For rainfed conditions at Mead, the models showed large

424   variability (uncertainty) in daily ETs from about -10 to +10 DAP (soil E dominated; Fig. 2)

425   similar to the irrigated field (Fig. 1). The greatest deviations (or errors) occurred from about 70

426   to 95 DAP when there was little rainfall (Fig. 2a). The observed ET continued at close to 4

427   mm/d, whereas the models simulated much lower rates. Like the irrigated field (Fig. 1), after

428   DAP 120 as the crop matured, the measured ET decreased rapidly, whereas the models continued

429   to simulate much higher ET. The ET variability during the -10 to +10 DAP period was related to

430   highly different methods for simulatingE, some of which proved to be less accurate. The issues

431   during the maturation period after 120 DAP are related to the insufficient termination of T after

432   maturity. More importantly, the period from about 70 to 95 DAP and beyond corresponds to the

433   period of water limitation, when most models (and the median) simulated lower than observed

434   ET. We suspect this is caused by inadequate soil water dynamics in the models, such as

435   insufficient rooting depth, inadequate water up-flux or the presence of a perched water table, as

436   well as excessive simulated ET during the early growth phase that depleted the simulated soil

437   water too much, thus reducing ET later.

438

439

440

Mead Irrigated NE2 2003

a. Weather

Tmax
Tmin
Tdew
Solar
Wind
Rain+Irr

b. Phase 2

Box Plots
Median
Observed

c. Phase 3

d. Phase 4

e. Observed and Medians for All Phases

Observed
Phase 2
Phase 3
Phase 4

Temperature (C), Wind (m/s), Solar (MJ m^{-2} d^{-1}), Rain+Irr (mm)

Daily Evapotranspiration (mm/day)

Days After Planting

441

442

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

458
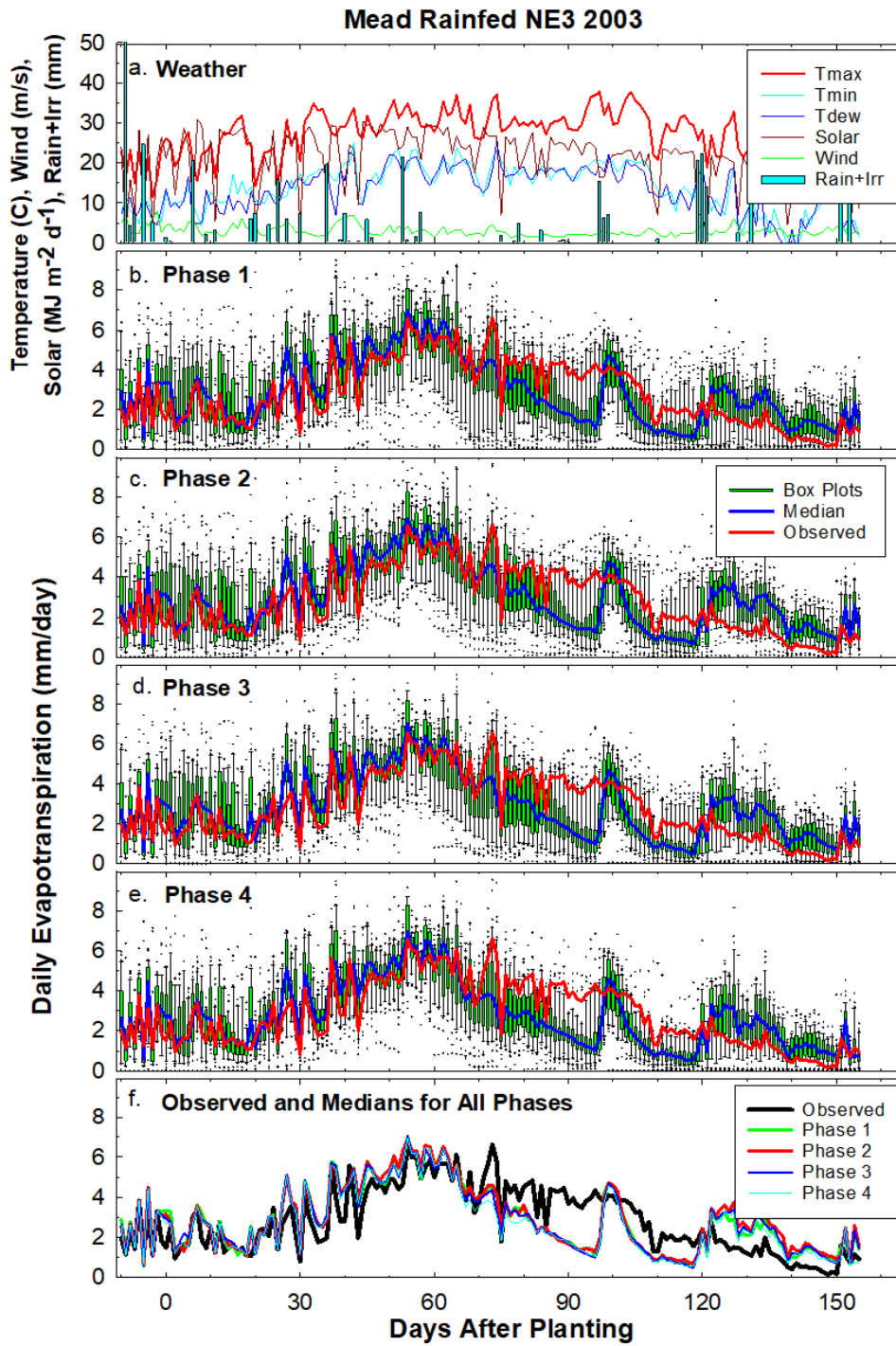
Figure 2. Similar to Figure 1 except for rainfed Mead field NE3 in 2003, and data for Phase 1 are also included.

464   3.1.1.3 *100% and 75% irrigated Bushland in 2013.* For the Bushland location, most of the

465   models (and the median) under-estimated ET during the 45 to 80 DAP period when windspeeds

466   were high (> 5 m/s) and dew points were low (Figs. 3, 4). Model calibration (Phases 1 to 4) only

467   partially improved this situation. This is possibly related to the fact that many of the models do

468   not adequately account for varying wind speed and humidity, as can be deduced from the fact

469   that the models estimated ET fairly well during periods of smaller ET but under-estimated ET

470   greatly during periods of larger ET, when wind speeds were high and relative humidity was low.

471   The fact that solar irradiance was also smaller during some of the periods of smaller ET (due to

472   storm fronts) indicates that the radiation and energy balance algorithms may also need

473   improvement. As before with Mead (Figs. 1, 2), the models failed to reduce T sufficiently after

474   crop maturation (Figs. 3, 4; 105 to 145 DAP). Surprisingly, the models tended to simulate the

475   75% irrigation treatment (Fig. 4) better than they did the 100% treatment (Fig. 3). Again, we

476   speculate that this is because many of the models had not been calibrated previously to account

477   for the very high winds and low humidity in Bushland, so their ETs simulations were lower than

478   the high observed ET rates for the 100% irrigations treatment (Fig. 3), whereas under the 75%

479   treatment (Fig. 4), drought stress reduced observed ET rates into the ranges for which the models

480   had been calibrated. The fact that observed ET for the 75% MESA irrigations treatment was

481   similar to that for the 100% SDI treatment (Evett et al., 2019) indicates that E may play an

482   important role in the discrepancies between simulated and observed ET for the 100% MESA

483   treatment. The major difference between SDI and MESA irrigation in the Bushland experiments

484   was the larger evaporative losses from the soil surface in MESA irrigated fields (Evett et al.,
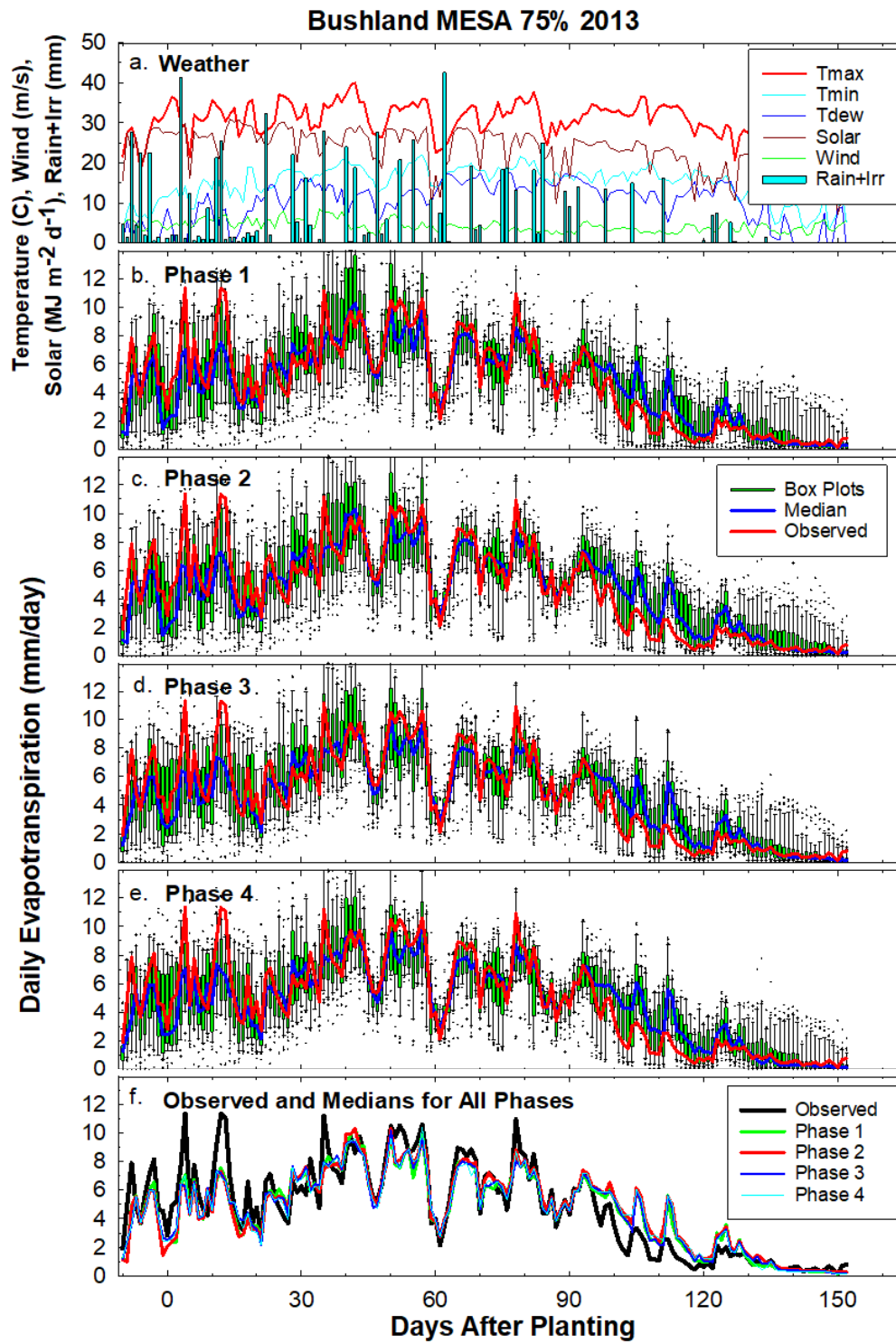
485   2019).

486
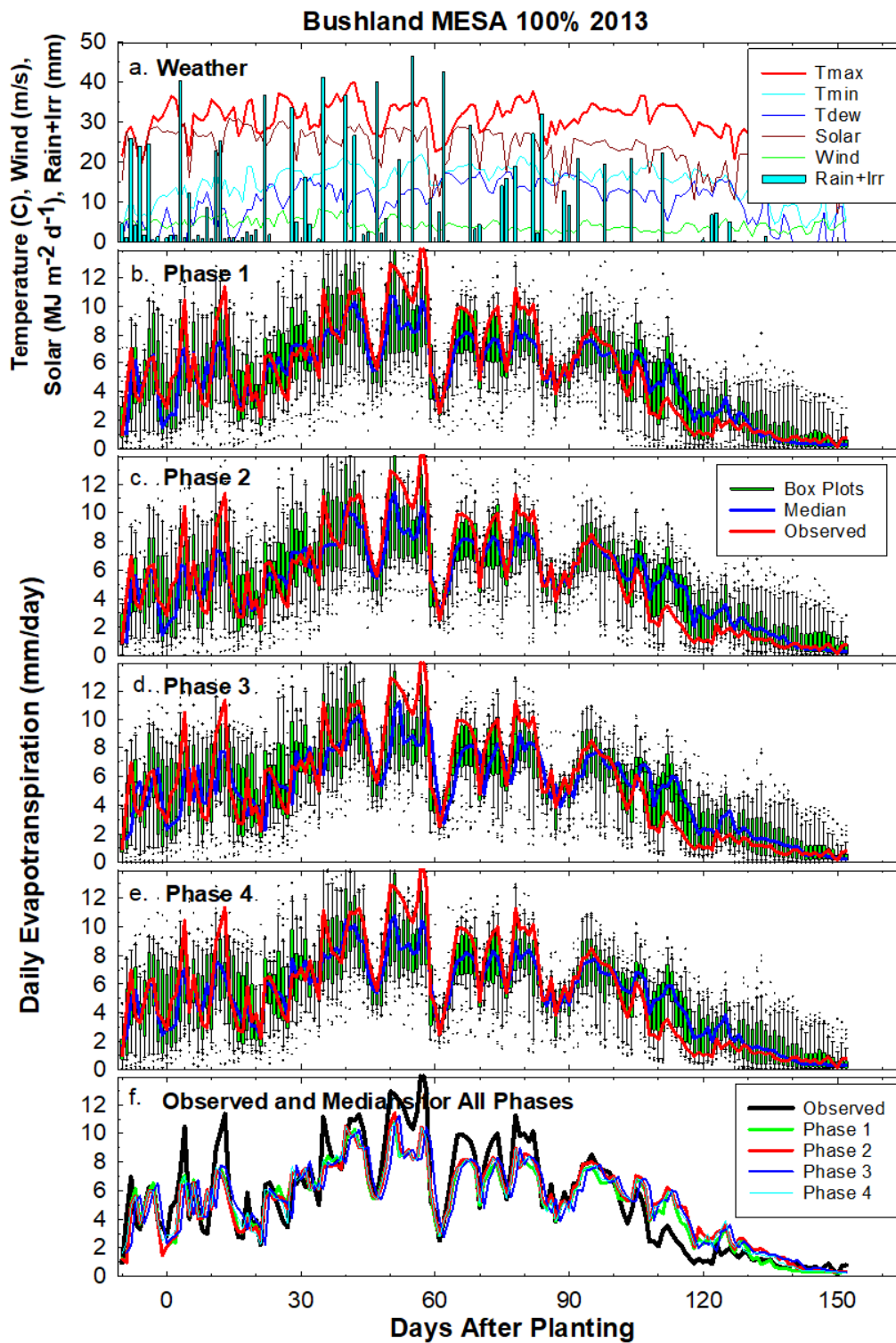
Figure 3. Similar to Fig. 1 except for 100% MESA (mid-elevation sprinkler application) irrigation at Bushland in 2013.

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

495　3.1.2 *Ranking of models with respect to their nRMSE for simulating daily ETs*

496　3.1.2.1 *Irrigated Mead in 2003*

497　The median of all the models had the lowest nRMSE for ETs for Phases 2, 3, and 4 for both early

498　season (-10 to +20 DAP; soil E dominant) and mid-season (41 to 100 DAP; canopy T dominate)

499　(Fig. 5). For early season STCK was the best model followed by several DSSAT "flavors," and

500　at mid-season several DSSAT flavors again did well, especially for Phase 2. STCK uses Penman

501　(1948) to calculate atmospheric demand and the 2-phase model of Brisson and Perrier (1991) and

502　Brisson et al. (1998, 2003) to calculate soil water evaporation, Ea (Table S1). Note that all the

503　DSSAT flavors listed for -10 to +20 DAP end in "R", which indicates that the soil E method of

504　Ritchie (1972) was better than the more recent method of Suleiman and Ritchie (2003, 2004).

505　However, for Phase 2 during the 41 to 100 DAP period DIFS and DCFS did well, but during this

506　period canopy T was dominant, so soil E was relatively unimportant then.

507

508　The effects of changes made by the modelers going from phase to phase can also be seen in Fig.

509　5. For example, BIOM was ranked 19th for Phase 2, -10 to +20 DAP but improved to 5th and 6th

510　for Phase 3 and Phase 4, respectively. Like the well-performing DSSAT flavors, BIOM

ETs Mead Irrigated 2003

511
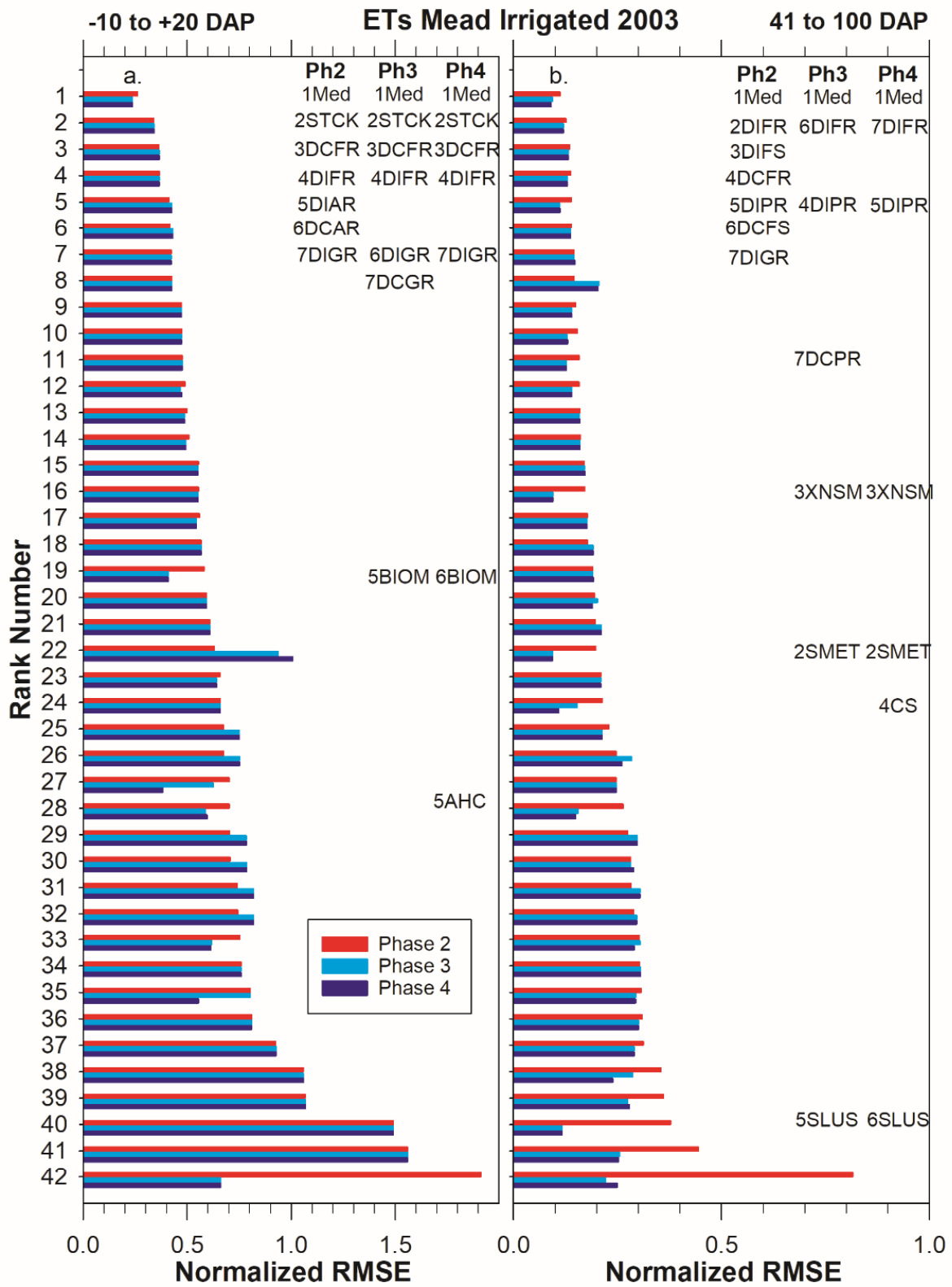
512 Figure 5. (a) Normalized root mean squared error (nRMSE) between observed and simulated
513 daily ET values from -10 to +20 days after planting (DAP)(mostly soil E) for the irrigated field
514 NE2 at Mead in 2003 for all the models. Phases 2, 3, and 4 are identified by red, cyan, and blue
515 bars with Phase 2 at the top and Phase 4 at the bottom of each group. Phase 1 data are missing
516 from this graph because a plant population mistake was made for Mead irrigated fields. The
517 models have been sorted in ascending order of nRMSE for Phase 2 from top to bottom of the
518 graph with the rank numbers on the left axis indicating their ranking for Phase 2. The Median
519 (Med) and the six best models (lowest nRMSE) for Phase 2 are listed under "Ph2". Somewhat
520 similarly, the Median and six best models for Phases 3 and 4 are also listed under "Ph3" and
521 "Ph4", but because the modelers made different adjustments going from phase to phase, their
522 rank order changed, so the names along with their nRMSE rank are in different positions down
523 the graph. (b) Same as for (a) except the data are for 41 to 100 DAP (mostly crop canopy T) with
524 the ranking done on the 41 to 100 DAP Phase 2 data

525

526   also uses Ritchie (1972) to simulate soil E. AHC rose from 28th to 5th from Phase 2 to Phase 4

527   for the -10 to +20 period. AHC uses the two-stage FAO-56 method to simulate E for mostly bare

528   soil (Table S1). A huge improvement was made by SLUS going from 40th for Phase 2 to 5th for

529   Phase 3 for the 41 to 100 DAP period. SLUS calculates atmospheric demand from Priestly and

530   Taylor (1972) and then uses an empirical equation to simulate potential ETp (Table S1), which

531   would be mostly T for the irrigated full canopy. XNSM, SMET, and CS all markedly improved

532   from Phase 2 to Phase 4 to be among the best for the full canopy (Fig. 5b). All three use FAO-56

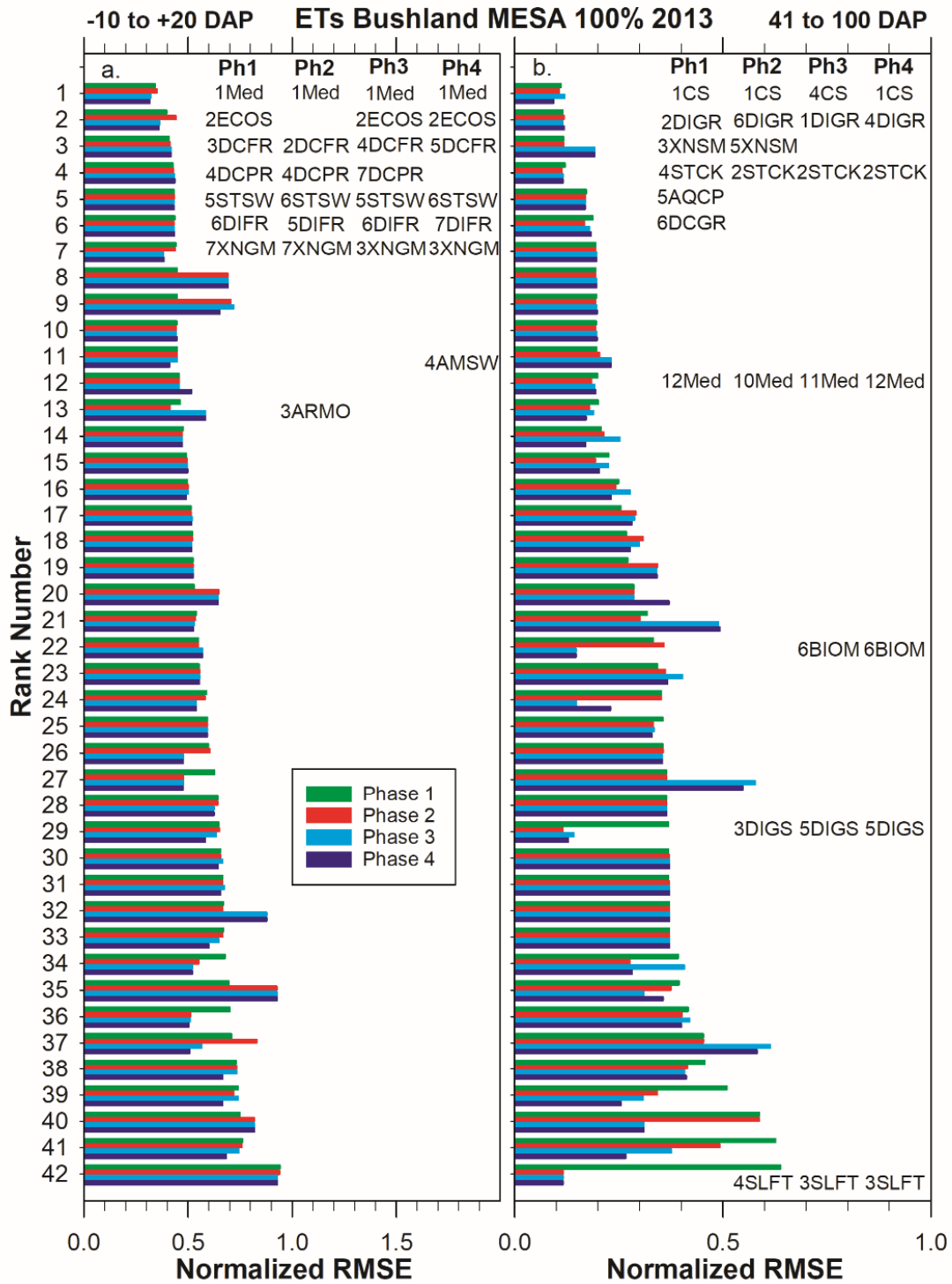533   (Allen et al., 1998) with some modifications (Table S1).

534

535   3.1.2.2 *Rainfed Mead in 2003*

536   The STCK model was best for simulating ETs for the -10 to +20 DAP period in the rainfed field

537   at Mead in 2003 for Phases 1 and 2, while the median was 2nd, and then they traded rankings for

538   Phases 3 and 4 (Fig. 6a). ECOS, JUL, DCFR, DIFR, and STSW also did very well. ECOS is a

539   full energy balance model while JUL uses the Penman-Monteith approach (Monteith, 1965) with

540   a 10-layer canopy (Table S1). BIOM rose from 31st for Phase 1 to 3rd for Phases 3 and 4. JUL

541   and XNGM were best for the 41 to 100 DAP period (Fig. 6b). MZD was 3rd for Phase 1, but did

542   much worse in the other phases. DIFR, DCFR, and DIPR did well. ECOS rose from 13th for

543   Phase 1 to 2nd for Phases 3 and 4. All of these listed models were better than the median for this

544   case.

545

546   3.1.2.3 *100% MESA irrigation at Bushland in 2013*

547   The median of all the models ranked 1st at simulating ETs from -10 to +20 DAP for all phases in

548   Bushland with 100% MESA irrigation in 2013 (Fig. 7a). Except for Phase 2, ECOS, an energy

31

549

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

561

562

Figure 7. Like Fig. 6 except for 100% MESA (mid-elevation sprinkler application) irrigation to
restore soil water to field capacity at Bushland.
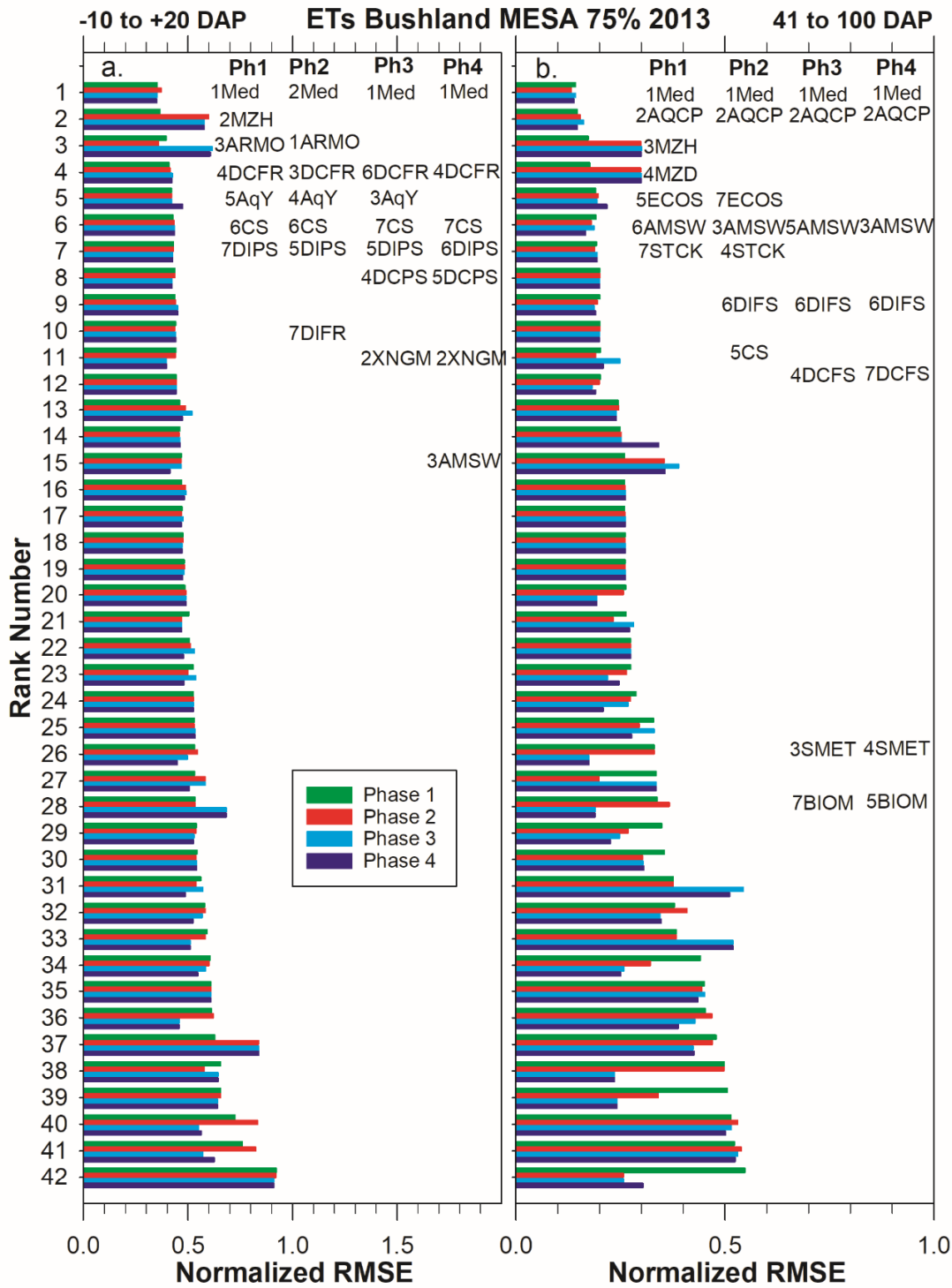
Figure 8. Like Fig. 6 except for the MESA (mid-elevation sprinkler application) 75% irrigation at Bushland.

570   balance model was best. DCFR, DCPR, STSW, DIFR, and XNGM all did well. At mid-season

571   (41 to 100 DAP, Fig. 7b), CS, DIGR, and STCK did well for all phases. BIOM and DIGS

572   improved greatly for phases 3-4. However, the median was only about 12th.

573

574   3.1.2.4 *75% MESA irrigation at Bushland in 2013*

575   The median of all the models was 1st for all but one phase for both early season (-10 to +20

576   DAP) and mid-season (41 to 100 DAP) for the 75% irrigation treatment at Bushland in 2013

577   (Fig. 8). MZH was ranked 2nd for Phase 1, early season (Fig. 8a) but then did much worse for

578   other phases. Similarly, ARMO did well for Phases 1 and 2, but then did much worse. DCFR,

579   CS, and DIPS did well in all phases. XNGM, AMSW, and DCPS were among the best for Phase

580   4. AMSW uses a transpiration efficiency to compute ETs from biomass accumulation, XNGM

581   uses a modified Penman-Monteith (Monteith (1965), and DCPS uses Priestly and Taylor (1972)

582   to simulate potential atmospheric demand and ultimately ETs.

583

584   At mid-season, AQCP and AMSW did well for all phases (Fig. 8b). MZH and MZD did well for

585   Phase 1, but then much worse for later phases. DIFS, DCFS, SMET, and BIOM did well for

586   Phases 3 and 4.

587

588   3.1.2.5 *Intercomparison among the models for all four cases of daily ET for Phase 4*

589   Looking at Figs. 5-8, no single model appears among the best (lowest nRMSE) six for all four

590   cases. The median was among the best for the all four cases from -10 to +20 DAP (mostly E),

591   but only for two cases from 41 to 100 DAP (mostly T). Focusing on the -10 to +20 periods

592   (mostly E), DCFR was among the best for 3 cases; STCK, DIFR, BIOM, ECOS, STSW, SNGM,

593 and AMSW for 2 cases; and DIGR, JUL, TMOD, CS, DIPS, DCPS for 1 case. For the 41 to 100

594 DAP periods, the median was among the best only twice. BIOM was best for 3 cases; DIFR, CS,

595 and SMET were best for 2 cases; and STCK DIGR, ECOS, JUL, STSW, XNGM, DIPR, XNSM,

596 SLUS, DIGS, SLFT, AQCP, AMSW, DIFS, and DCFS were all among the best for 1 case.

597 BIOM stands out as being the only model to be among the best twice for early season (mostly E)

598 and thrice for midseason (mostly T).

599

600 3.2 *Inter-comparisons within the DSSAT family*

601 3.2.1 *Daily ETs*

602 A comparison of E methods within the DSSAT models, revealed that the older Ritchie-2-stage

603 model (Ritchie, 1972) was consistently better (lower nRMSE and lower simulated ETs) than the

604 Sulieman and Ritchie method (2003, 2004) during the -10 to +20 DAP period, regardless of the

605 other ET methods (Figs. 9a, 10a). The Ritchie-2-stage method was also better (slightly lower

606 nRMSE) for ETs in the 41 to 100 DAP full canopy phase (Figs. 9b, 10b) for two reasons (less E

607 during that phase, but mostly because lower early E allowed soil water in deeper layers to be

608 conserved for the 41 to 100 DAP period, thus contributing more to T during the latter phase).

609

610 In spite of having a theoretically more realistic mechanism for moving soil water with potential

611 gradients, the Hydrus method (ŠŠŠimůnek et al., 1998, 2008; Shelia et al., 2018) did not perform

612 as well as the more empirical Ritchie (1972) and Sulieman and Ritchie (2003, 2004) methods

613 (Fig. 9a, 9b). However, Hydrus was just recently incorporated into the DSSAT shell, whereas the

614 Ritchie (1972) and the Sulieman and Ritchie (2003, 2004) routines have been used for many

615 years and likely have been fine-tuned to the system. Also, Hydrus is very sensitive to the values

616   of the soil physical and hydraulic properties, so if those parameter values were off, the simulated

617   ET would also be off.

618

619   A comparison of potential ET (ETp) methods within the DSSAT models illustrated that the

620   FAO-56 method (in present DSSAT; Allen et al., 1998) with Kcan of 0.62 (gives Kep = 0.50)

621   performed better (lower nRMSE) for ETs than the other ETp methods: Priestley-Taylor (P-T;

622   1972), alfalfa reference-[ET_r, ASCE equation (Allen et al., 2005)], or grass reference-[ET_o,

623   ASCE equation (Allen et al., 2005)] during both the -10 to +20 DAP period and the 41-100 DAP

624   period (Figs. 10a, 10b). Kcan is the extinction coefficient for absorption of photosynthetically-

625   active radiation by LAI, while Kep is the extinction coefficient for absorption of total solar

626   energy by LAI. The default Kcan for CERES is 0.85 (in the ecotype file). Kcan was reduced to

627   0.62
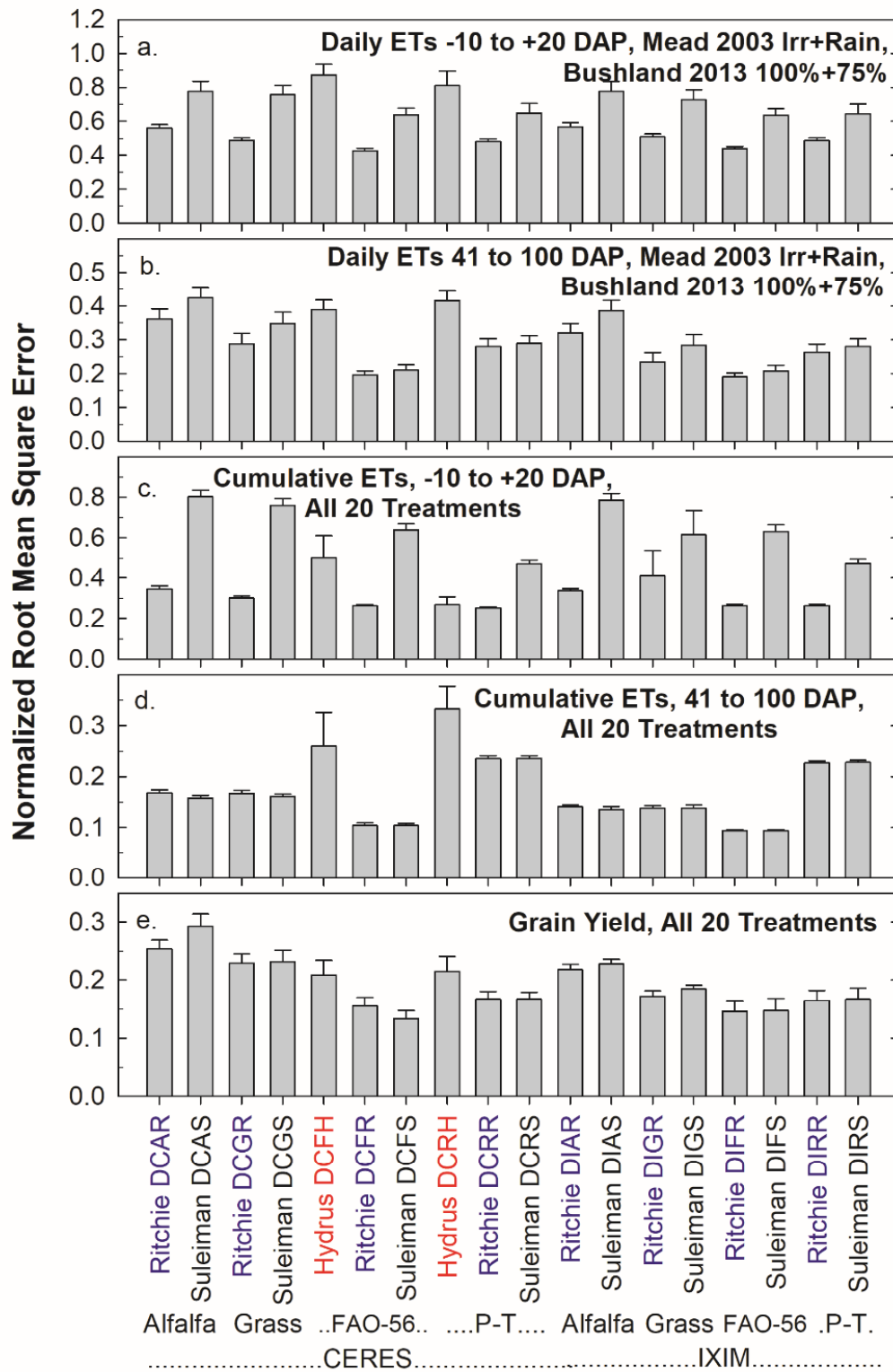
630   Figure 9. Normalized root mean square errors (nRMSE) of the 18 "flavors" of the DSSAT family
631   models (a) for the -10 to +20 DAP periods (mostly soil E) of daily ETs over all phases for the
632   irrigated and rainfed data for Mead 2003 and the 100% and 75% MESA irrigated data for
633   Bushland 2013. Models included are DSSAT CSM-CERES and DSSAT CSM-IXIM, whose
634   horizontal names span the corresponding left ten and right eight vertical bars, respectively.
635   Potential ETp calculation methods are using alfalfa (tall, ET_r) and grass (short, ET_o) reference
636   crop coefficients with the ASCE standardized reference equation (Allen et al., 2005), FAO-56
637   (Allen et al., 1998), and Priestley-Taylor (1972). These horizonal names span the corresponding
638   bars above them. Soil evaporation calculation methods follow Ritchie (1972; labelled
639   "Ritchie"), Suleiman and Ritchie (2003, 2004; labelled just "Suleiman"), and Hydrus (Šimůnek
640   et al., 1998, 2008; Shelia et al., 2018; labelled "Hydrus"). (b) Like (a) except for the 41 to 100
641   DAP periods. (c.) The values plotted are averages (+ standard errors) of the nRMSEs for Phase 4
642   for all 20 treatment-years of the cumulative ETs from -10 to +20 DAP periods. (d.) Like (c.)
643   except for the cumulative ETs from 41 to 100 DAP. (e.) nRMSEs for Phase 4 grain yields for all
644   20 treatments.
645

646
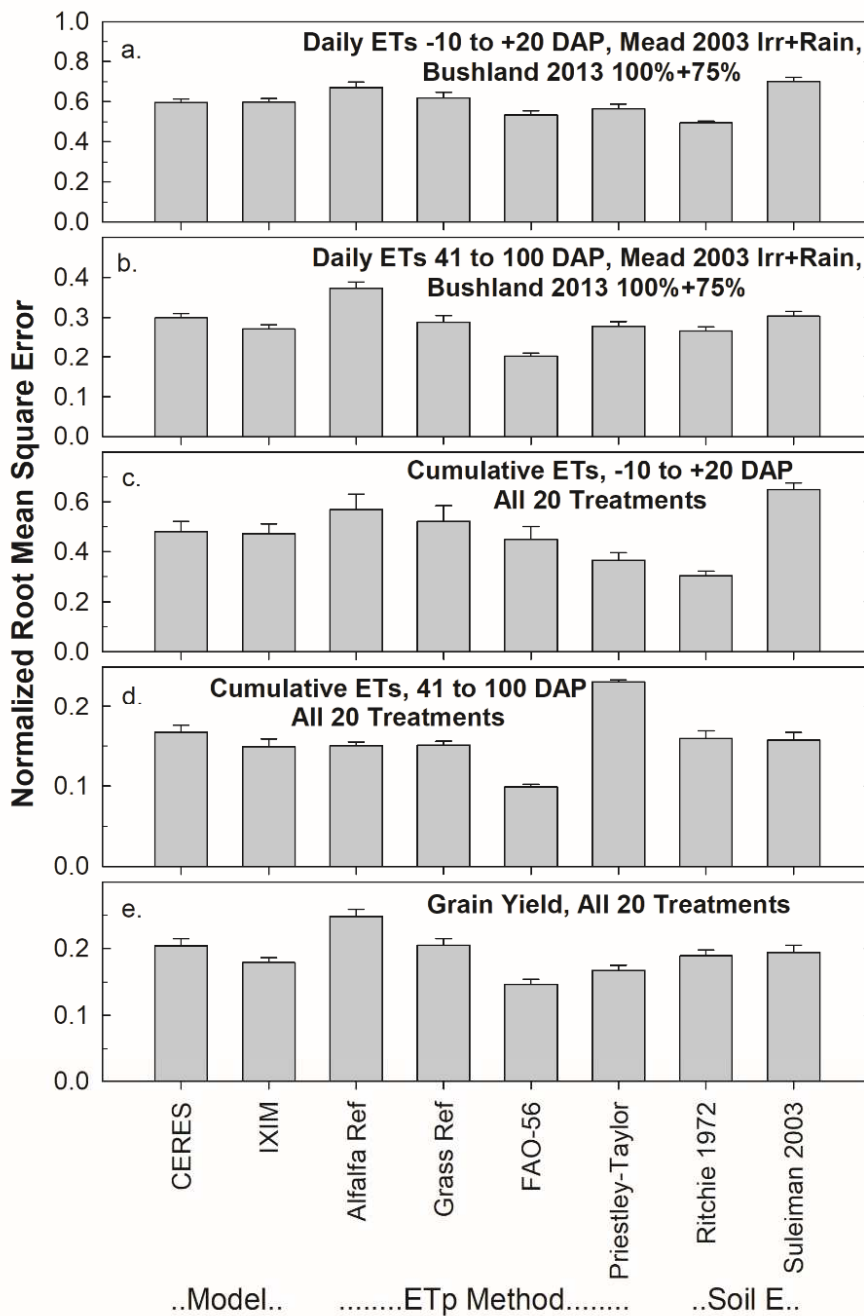
648
649   Figure 10. Direct comparisons using the same data as for Fig. 9 (excluding Hydrus) between the
650   DSSAT-CERES and DSSAT-IXIM models, among the four potential ET methods, and between
651   the two soil water evaporation methods for the corresponding a, b, c, d, and e graphs. The
652   horizontal "Model", "ETp Method", and "Soil E" labels span the corresponding bars above.
653
654

655 during phase 3, which reduces the effective energy extinction from 0.685 to 0.50 [latter value

656 supported lysimeter studies of Villalobos and Fereres (1990), as well as the theory of foliar

657 absorption of total solar energy (Goudriaan, 1977)]. The Kep=0.50 was used for P-T as well. On

658 the other hand, the alfalfa reference-FAO-56, or grass reference-FAO-56 are dual-coefficient

659 methods that compute their own coefficients during incomplete and full canopy phases of ET. In

660 contrast to a previous study on cotton (*Gossypium hirsutum* L.) ET (Thorp et al., 2020), the

661 methods based on ASCE alfalfa and grass reference ET did not perform as well as DSSAT FAO-

662 56 and P-T; however, the calibration methodology limited their comparability in the present

663 study. It appears that the newly reduced Kep of 0.50 contributed to improved DSSAT

664 performance, and it is an improvement over the default DSSAT value. As mentioned previously,

665 Sau et al. (2004) reported that the FAO-56 with a Kep=0.50 gave the best simulations of ET, soil

666 water extraction, and biomass accumulation with the CROPGRO-Faba bean model for a water-

667 limited environment. FAO-56 was better than P-T, and the extinction coefficient (Kep=0.50) was

668 better than a higher Kep for either ET method. Similarly, Lopez-Cedron et al. (2008) found the

669 CERES model gave better simulations of maize biomass, grain yield, and harvest index under

670 water-limited environments, using FAO-56 rather than P-T, and again, Kep of 0.50 was better

671 than a higher energy extinction coefficient (default in CERES was 0,685).

672

673 There was no significant difference in nRMSE between the CERES or IXIM models for the -10

674 to +20 DAP period (Fig. 10a, soil E dominant), whereas for the 41 to 100 DAP period (Fig. 10b,

675 canopy T dominated), IXIM was slightly better, likely because of its more realistic simulation of

676 LAI progression. IXIM senesces green leaf area more rapidly (and more mechanistically) near

677   maturity than does CERES, which results in less T during the grain-filling phase, and which

678   more closely matches the observed reduction inT.

679

680   Comparing methods for calculating potential evapotranspiration (ETp) on the nRMSE of ETs for

681   the 41 to 100 days after planting (DAP) period (Fig. 10b), the FAO-56 method had significantly

682   lower *nRMSE. For the -10 to +20 DAP period (Fig. 10a), it was better than both the alfalfa (*tall;

683   *Medicago sativa* L.) and grass (short) crop coefficients with the ASCE standardized reference

684   equation, but Priestley-Taylor (P-T) tended to be almost as good. Comparing soil E methods,

685   Ritchie (1972) was much better than Suleiman and Ritchie (2003, 2004) for the -10 to +20 DAP

686   period (Fig. 10a) soil E dominant), and Ritchie (1972) was slightly better even for the 41 to 100

687   DAP period (Fig. 10b, canopy T dominant).

688

689   *3.3 Inter-comparisons within the STICS, Expert-N, and MAIZSIM families*

690   Comparing the nRMSE of daily ETs between the two "flavors" of each pair of the STICS,

691   Expert-N, and MAIZSIM families, there were no significant differences (Figs. 11a, 11b). STCK

692   uses a single surface model (Penman, 1948) to compute potential ETp, whereas STSW handles

693   separate canopy and soil surfaces (Shuttleworth and Wallace, 1985), Thus, for these four cases,

694   STCK and STSW performed equally well at simulating soil E (Fig. 11a) and canopy T (Fig. 11b)

695   in spite of the different methods for simulating ETp. Both XNSM and XNGM models use

696   Penman-Monteith based approaches for simulating ETp. However, XNSM follows FAO 56

697   guideline based on ETo multiplied with a single crop factor to get ETp while in XNGM the

698   required surface- and aerodynamic resistances are calculated directly from simulated LAI and

699   simulated canopy height. In addition, XNGM follows the more detailed Farquhar model in

700   simulating photosynthesis and leaf T but simplifies vertical root distribution. The latter could

701   possibly explain slightly better soil moisture simulations of XNSM compared to XNGM (data

702   not shown). In XNSM, temperature, moisture, and nutrient availability in different soil layers are

703   taken into account when simulating rooting depth and root length distribution. In contrast,

704   XNGM assumes a uniform distribution of root length density within the rooted zone, with the

705   increase in rooting depth simply simulated from the increase in root biomass, regardless of the

706   soil conditions. Thus, considering that there are marked differences between the two models, it is

707   surprising that they differ so little in their ability to simulate ETs. The lack of significant

708   differences between MZD and MZH is reasonable because they are the same in their

709   representation of plant and soil processes. Both models run on an hourly time step internally but

710   MZD takes daily weather data as input and interpolates them into hourly time steps, while MZH

711   takes hourly weather data directly as input.

712

713   3.4 *Potential ETp and other sources of variability/error in daily ETs*

714   There was a wide variability among the models in their simulated values for daily ETs as shown

715   in Figs. 1-4, which is similar to the previous results reported by Kimball et al. (2019). In that

716   report, Fig. 10 shows that much of the variability can be attributed to variability among the

717   models in their values of ETp. Therefore, for this study we requested more values of "upstream"

718   variables that the modelers might be using to compute ETp, including reference ET based on

719   short (12 cm) grass (ETo), reference ET based on tall (50 cm) alfalfa (ETr), soil coefficient (Ks),

720   basal crop coefficient (Kcb), soil evaporation coefficient for drying soil (Ke), overall crop

721   coefficient (Kc), potential soil evaporation (Ep), potential transpiration (Tp), ETp, and of course,

722
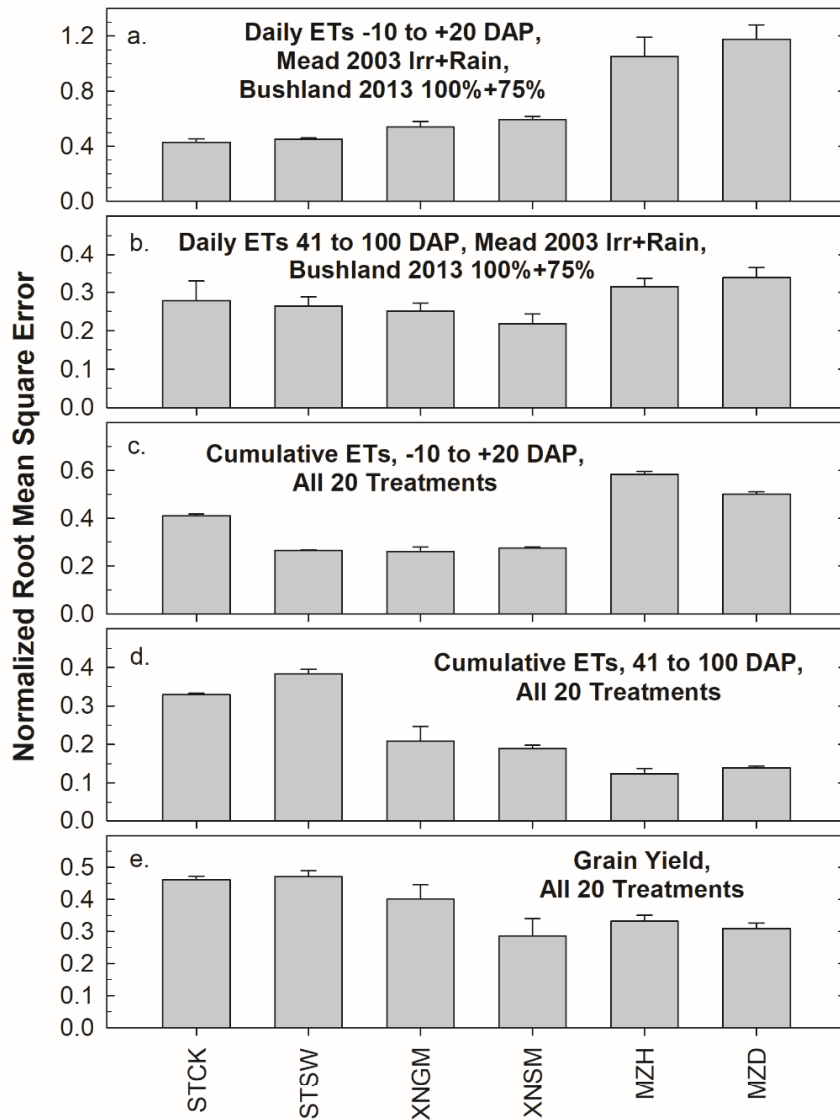
Figure 11. Direct comparisons using nRMSE between the STCK and STSW flavors of the
STICS model family, between XNGM and XNSM flavors of Expert-N family, and between the
MZH and MZD flavors of the MAIZSIM model for (a) the -10 to +20 DAP time period (mostly
soil E). The data used were all phases for the irrigated and rainfed data for Mead 2003 and the
100% and 75% MESA irrigated data for Bushland 2013. (b). Like (a) but for the 41 to 100 DAP
period (mostly canopy T). (c.) Phase 4 of cumulative ETs from -10 to +20 DAP for all 20
treatments. (d) Like (c) but for cumulative ETs from 41 to 100 DAP. (e) Phase 4 grain yield for
all 20 treatment-years.

44

731 ETs. Three of the models did not report ETp, presumably the energy balance ones that do not use

732 the concept.

733 Focusing on the Phase 2 results from irrigated Mead in 2003, 34 models reported Ep and 35

734 models reported Tp, and both Ep and Tp were quite variable (data not shown), As expected, the

735 magnitude and variability of the soil Ep were greatest for bare soil at the beginning of the season.

736 However, there was more than a 2 mm/day spread even at mid-season. Surprisingly, a few of the

737 models showed some Tp starting on the day of planting before the plants had even emerged.

738 Then, as the Tp increased in magnitude as plants grew to full size by mid-season, so did the

739 range in variability among them, similar to ETs.

740

741 Thirteen of the models reported ETo and only 4 reported ETr. Presumably ETo and ETr depend

742 only on weather, yet ETo varied by a factor of about 2 at midseason among the 13 models (data

743 not shown). Apparently, several different definitions and equations for ETo are in play among

744 these models.

745

746 Only 6, 4, 4, and 7 models used Ks, Kcb, Ke, and Kc, respectively. It seems likely that more

747 models do use them, but they are computed and not routine output, so the modelers would have

748 had to change code to get them. In any event, there appear to be several ways that models are

749 getting from ETo (or ETr) to ETp that are contributing to the variability of ETs.

750

751 Thus, in conclusion, the variability in ETp and ETs appears to be coming from steps all along the

752 way starting from the calculations of ETp to the final resultant ETs.

753

754 *3.5 Cumulative month to whole season ETs results for all 20 treatment-years*

755 The previous sections focused on the daily ET for four selected treatment-years. However, one

756 can imagine that an underestimate of simulated daily ET one day could save some simulated soil

757 moisture and lead to an overestimate the next day. The following sections examine the

758 cumulative ET over longer time periods to reveal the extent that the errors are also cumulative.

759

760 *3.5.1 ETs from -10 to +20 DAP (mostly soil E) and 41 to 100 DAP (mostly canopy T)*

761 Moving from daily ETs for the four cases (2003 for irrigated and rainfed Mead; 2013 MESA at

762 75% and 100% irrigation at Bushland) to cumulative ETs over longer time durations for all 20

763 treatment-years also showed wide variability among the models (Fig. 12). Again, there were

764 variations by factors of 2 to more than 4 among them in cumulative ETs from -10 to +20 DAP

765 (mostly soil E) (Fig. 12a). There was little or no improvement in going from Phase 1 to Phase 4.

766 For Treatments 1-10 for Mead, the medians of the models were close to the observations, but for

767 Treatments 11 and 12, the models generally overestimated ETs. For Bushland, most of the

768 models underestimated Treatment 13 when spray irrigation wetted the surface and Treatment 18

769 when rainfall wetted the surface of SDI fields. Most models overestimated Treatments 17 and 19

770 when the SDI field surface was dry despite plentiful irrigation, but the medians were close to

771 observed for the other 4 treatments. These results indicate problems simulating E from wetted

772 surfaces and with simulated redistribution of water from buried drip lines to the surface (too

773 much water movement to the surface).

774

775 Looking at cumulative ETs from 41 to 100 DAP (mostly canopy T), there is a range of about a

776 factor of 2 among the models (Fig. 12b), which is bad but less than that from the bare soil (Fig.

46

777   12a). For Mead, most of the models overestimated ETs for Treatments 1-6 and 8-11. They

778   underestimated Treatment 7 but were close for Treatment 12. For Bushland, most of the models

779   underestimated ETs under sprinkler irrigation for Treatments 14-16, which represent wetter soil

780   Ranking the models' ability to simulate cumulative ETs from -10 to +20 DAP by nRMSE (Fig.

781   13a), the medians were close to observations for Phases 2-4. SLFT was the best model for Phases

782   2 and 3 and was next best in Phase 4. SLFT uses FAO-56 (Allen et al., 1998) to calculate

783   atmospheric demand and then dual crop coefficients simulate ETs (Table S1). For Phase 2,

784   models in the DSSAT family were ranked 3-7, and several did well in Phases 3 and 4. AMSW

785   was best in Phase 4. CS and XNGM were among the best in Phases 3 and 4.

786

787   Similarly ranking their ability to simulate ETs from 41 to 100 DAP (Fig. 13b), several of the

788   models in the DSSAT family did well for Phases 2, 3, and 4. ECOS was among the best for

789   Phases 2 and 3. SSMi (which uses Priestly and Taylor (1972) for potential atmospheric demand

790   and transpiration efficiency with biomass accumulation to simulate ETs) was ranked 6 for Phases

791   3 and 4, and SMET was ranked 3rd for Phase 4. surface conditions, but the medians were close

792   for Treatment 13. Under SDI irrigation, most models underestimated Treatment 18, but the

793   medians were close for Treatments 17, 19, and 20.

794

795
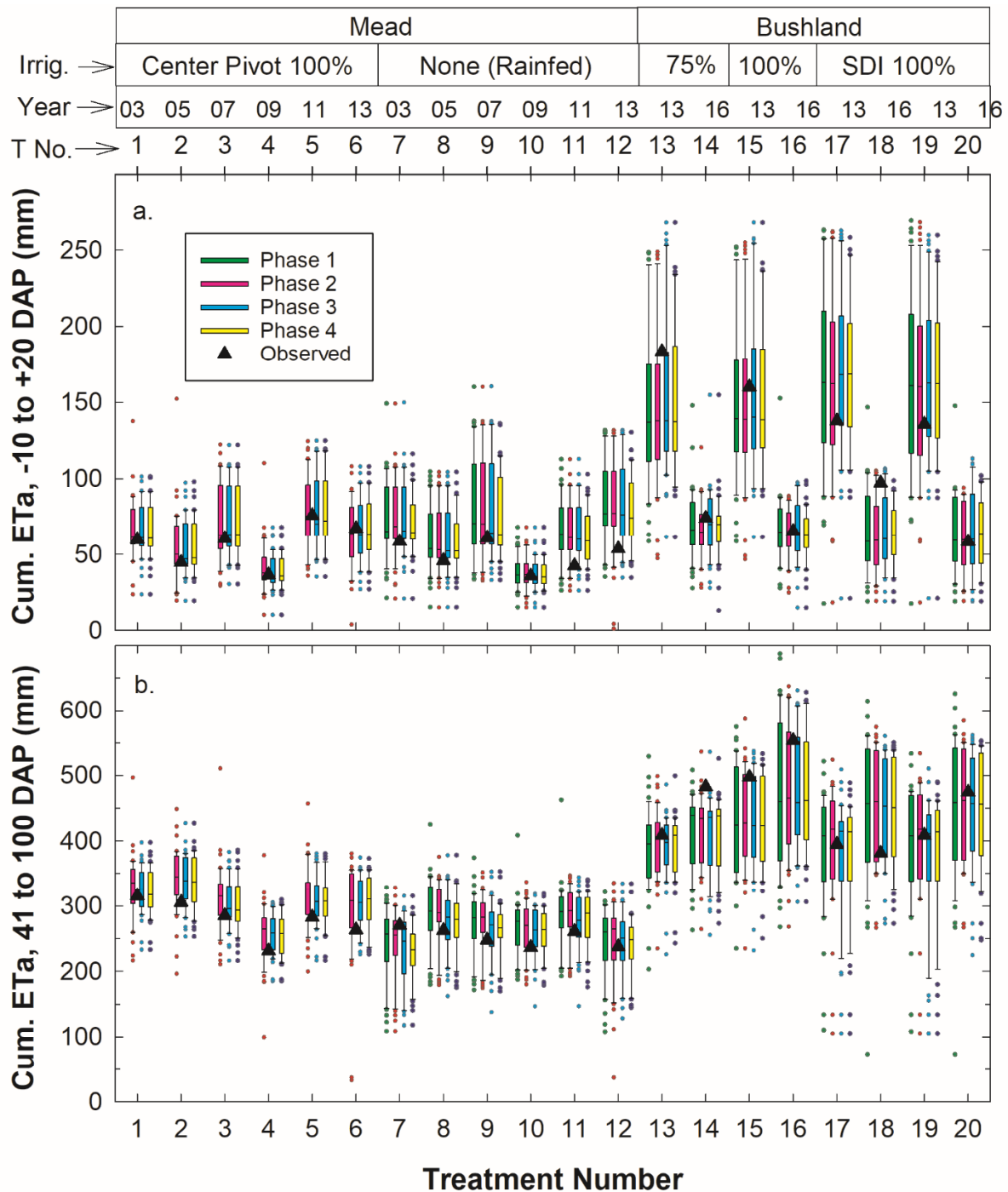
Figure 12. Box plots for all 20 treatment-years (as defined at the top) of cumulative simulated evapotranspiration (ETs) over (a) the -10 to +20 days after planting (DAP) time period (mostly E) and (b) the 41 to 100 DAP time period (mostly T) for all four phases. The dark lines across the boxes indicate the medians of all the models. Also shown are the corresponding observations. Phase 1 is not shown for treatments 1-6 because of a planting density mistake.

48

803   Looking back at Section 3.1.2.5, BIOM was among the best at simulating daily ETs, yet it was

804   not among the best at simulating ETs over the longer intervals. On the other hand, DIFR was

805   almost as good as BIOM for simulating daily ETs, and it was best for simulating cumulative ETs

806   over the 41 to 100 DAP periods (Fig. 13, Phase 4). Besides DIFR, DCFS and SMET were the

807   only other two models that were among the best for cumulative ETs over the 41 to 100 DAP

808   periods and also were among the best for at least one case of daily ETs. For the -10 to +20 DAP

809   periods, DCFR, XNGM and CS are the only models that were best for simulating cumulative

810   ETs and also for daily ETs at least for one case. Thus, doing well for simulating daily ETs did

811   not guarantee success at simulating cumulative ETs.

812

813   3.5.2 *Inter-comparisons of cumulative ETs within the DSSAT and other model families*

814   There were wide differences in performance among model "flavors" within the DSSAT family

815   for cumulative simulated ETs from -10 to +20 DAP (mostly soil E) over the 20 treatment-years

816   (Fig. 9c). Most obvious is that the Ritchie (1972) soil E method did much better than the

817   corresponding Suleiman (Suleiman and Ritchie, 2003, 2004) method for every case. The Hydrus

818   method did comparatively well for this cumulative-ETs/20-treatment-year comparison, which is

819   in contrast to the daily-ETs/4-treatment-year comparison in Fig. 9a.

820

821

50
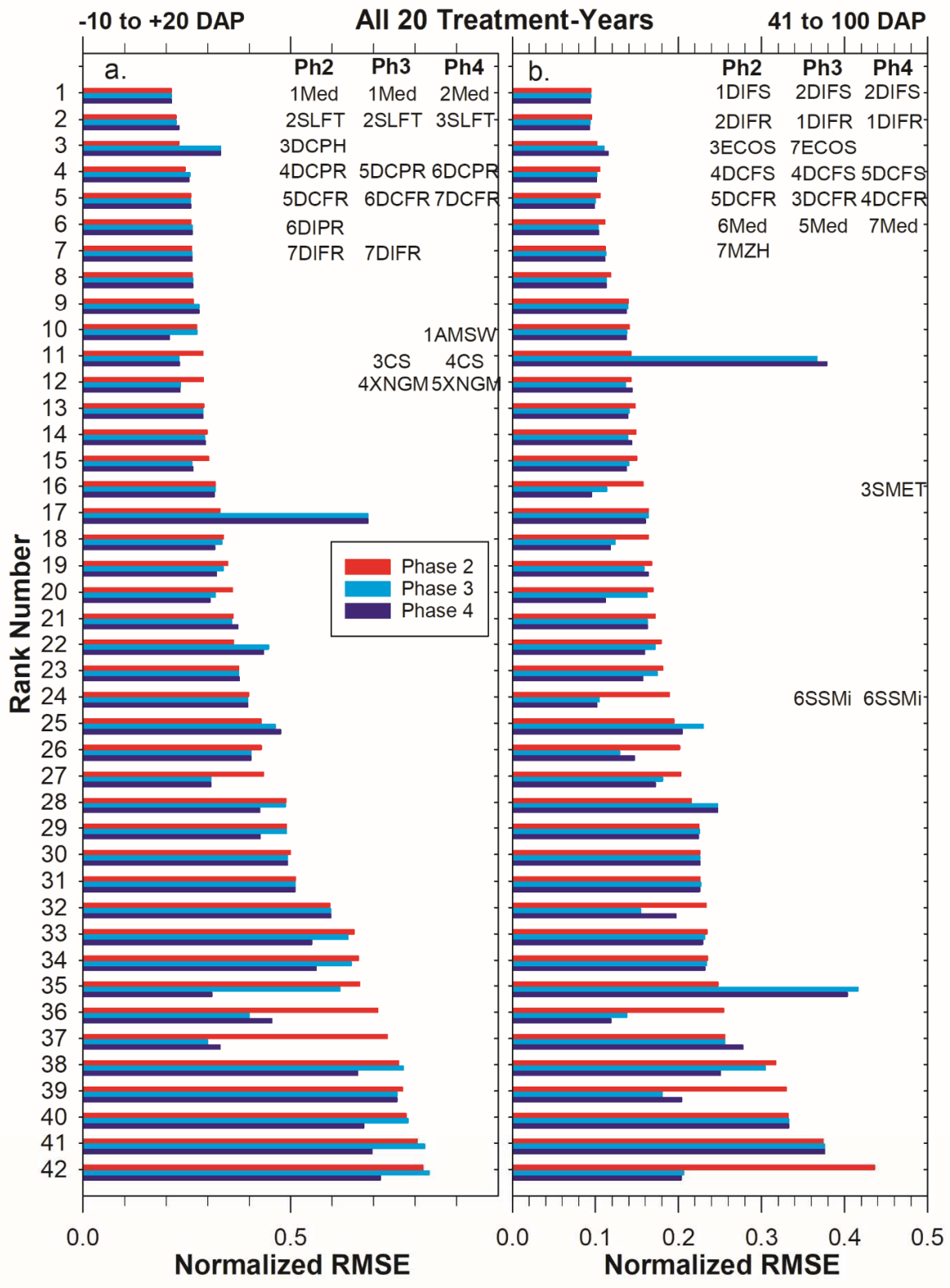
822   Figure 13. (a.) Normalized root mean squared error (nRMSE) between observed and simulated
823   cumulative ET values from -10 to +20 days after planting (DAP)(mostly soil E) for all 20
824   treatment-years for all the models. Phases 2, 3, and 4 are identified by red, cyan, and blue bars
825   with Phase 2 at the top and Phase 4 at the bottom of each group. The models have been sorted in
826   ascending order of nRMSE for Phase 2 from top to bottom of the graph with the rank numbers
827   on the left axis indicating their ranking for Phase 2. The Median (Med) and the six best models
828   (lowest nRMSE) for Phase 2 are listed under "Ph2". Somewhat similarly, the Median and six
829   best models for Phases 3 and 4 are also listed under "Ph3", and "Ph4", but because the modelers
830   made different adjustments going from phase to phase, their rank order changed, so the names
831   along with their nRMSE rank are in different positions down the graph. Phase 1 is not shown
832   because of the planting density error for the six irrigated Maize treatment-years. (b.) Same as for
833   (a.) except the data are for 41 to 100 DAP (mostly crop canopy T).
834

835 As would be expected, looking at the 41 to 100 DAP periods, the soil E method had little effect

836 (Fig. 9d). However, Hydrus, did poorly which is in contrast to the -10 to +20 periods (Fig. 9c).

837

838 There was no significant difference in performance between the CERES Maize and IXIM Maize

839 models for the -10 to +20 DAP periods (Fig. 10c), whereas IXIM was slightly better than

840 CERES for the 41 to 100 DAP periods (Fig. 10d). The better performance of IXIM for full

841 canopy conditions was likely because of its more realistic simulation of LAI progression, as

842 mentioned previously. Priestley-Taylor was the best ETp method for the -10 to +20 DAP periods

843 (Fig. 10c) but worst for the 41 to 100 DAP periods (Fig. 10d). FAO-56 was second best for -10

844 to +20 DAP periods (Fig. 10c) but best for the 41 to 100 DAP periods (Fig. 10d). As was

845 obvious from Fig. 9c, the direct comparison between Ritchie (1972) and Suleiman and Ritchie

846 (2003) in Fig. 10c, confirms the superiority of the older Ritchie (1972) method for simulating

847 soil E, likely because the Suleiman and Ritchie overestimates the upward movement of soil

848 water from deeper depths. However, under full canopy conditions (Fig. 10d), there was no

849 difference between the two soil E methods.

850

851 Looking at other models with more than one flavor, STSW performed better than STCK for

852 cumulative ETs from -10 to +20 DAP (Fig. 11c), but the reverse was true from 41 to 100 DAP

853 (Fig. 11d). It is somewhat surprising that the two-surface method for computing ETp in STSW

854 did better for the -10 to +20 DAP period when there was only the single soil surface, but was

855 worse for the 41 to 100 DAP full canopy period. However, looking more closely, both models

856 did well for both time periods for the Mead data, whereas for Bushland in 2013, both models had

857 trouble getting emergence with SDI in 2013, and this issue distorted the results. There was no

858   significant difference between XNGM and XNSM for either of the time periods (Figs. 11c, 11d).

859   As noted in section 3.3, the two models use slightly different variants of the Penman-Monteith

860   approach and differing root distribution approaches resulting in essentially no differences in

861   daily ETs for the four cases (Figs 11a, 11b) nor in cumulative ETs for all 20 treatments (Figs.

862   11c, 11d). MZD did slightly better than MZH for the -10 to +20 DAP periods (Fig. 11c), but

863   there was little difference for 41 to 100 DAP (Fig. 11d). Any differences between MZD and

864   MZH are likely associated with the differences between interpolated and measured hourly

865   weather data that were driving MZD and MZH, respectively.

866

867   3.6 *Ability of the models in Phase 4 to simulate agronomic parameters for all 20 treatment-years*

868   *– maximum leaf area index, biomass at about 40 DAP and about 100 DAP, and final*

869   *grain yield.*

870   3.6.1 *Considering all the models*

871   There was a wide range in simulations of maximum LAI between the lowest and the highest

872   models (Fig. 14a). However, for some treatments, most of the models agreed closely as indicated

873   by short boxes. Indeed, for Treatments 1 and 3, most of the models agreed almost exactly with

874   one another and with observations. For many treatments, the medians agreed closely with

875   observations. However, for Treatments 4 and 14, the models mostly underestimated LAI,

876   whereas for Treatments 11, 15, and 17, they mostly overestimated LAI. For Bushland, treatments

877   15, 17, and 19 were in the 2013 year that began quite dry and required plentiful irrigation to

878   achieve germination and to support crop growth. Overestimation of LAI may be linked to model

879   algorithms overreacting to the plentiful irrigation in an otherwise stressful year.

880

881   Most of the models overestimated above-ground biomass at about 40 DAP for almost all the

882   treatments (Fig. 14b). This was particularly true for the dry 2013 year at Bushland, again

883   indicating that the plentiful irrigation caused the models to overestimate biomass accumulation

884   despite an otherwise stressful environment. However, by 100 DAP (Fig. 14c), most of the

885   models did much better, and agreement with observations was much closer. For final grain yield,

886   most of the models did surprisingly well (Fig. 14d). For the irrigated Mead data (Treatments 1-

887   6), most of the models agreed with one another and with the observations. They also did well for

888   four of the Mead rainfed years, but underestimated Treatments 7 and 12. They did less well with

889   the Bushland data, especially underestimating the SDI irrigation grain yields. The

890   underestimation of SDI grain yields is likely tied to overly large partitioning of applied water to

891   soil E, leaving less available water for T and grain yield formation. Many models, including

892   DSSAT, lack true SDI capability and applied the water to the soil surface in this study. Because

893   SDI was more efficient in water use than the MESA irrigation method in the actual fields used

894   for this study (Evett et al., 2020) and, therefore, likely will be more widely used in the future, the

895   inability to handle SDI is an emerging lacuna in many of the models that should be addressed in

896   future.

897

898   3.6.2 *Inter-comparisons of grain yield within the DSSAT and other model families*

899   DCFS was the best of the several model flavors within the DSSAT family to simulate grain

900   yield, as indicated by nRMSE for Phase 4 (Fig. 9e). However, general patterns are not obvious in

901   Fig. 9e. Nevertheless, some patterns emerge from a direct comparison in Fig. 10e. IXIM was

902   slightly better than CERES. FAO-56 emerged as the best ETp method followed by Priestly-

903   Taylor and then ASCE standardized reference ET equation with grass (short, 12 cm) crop

904
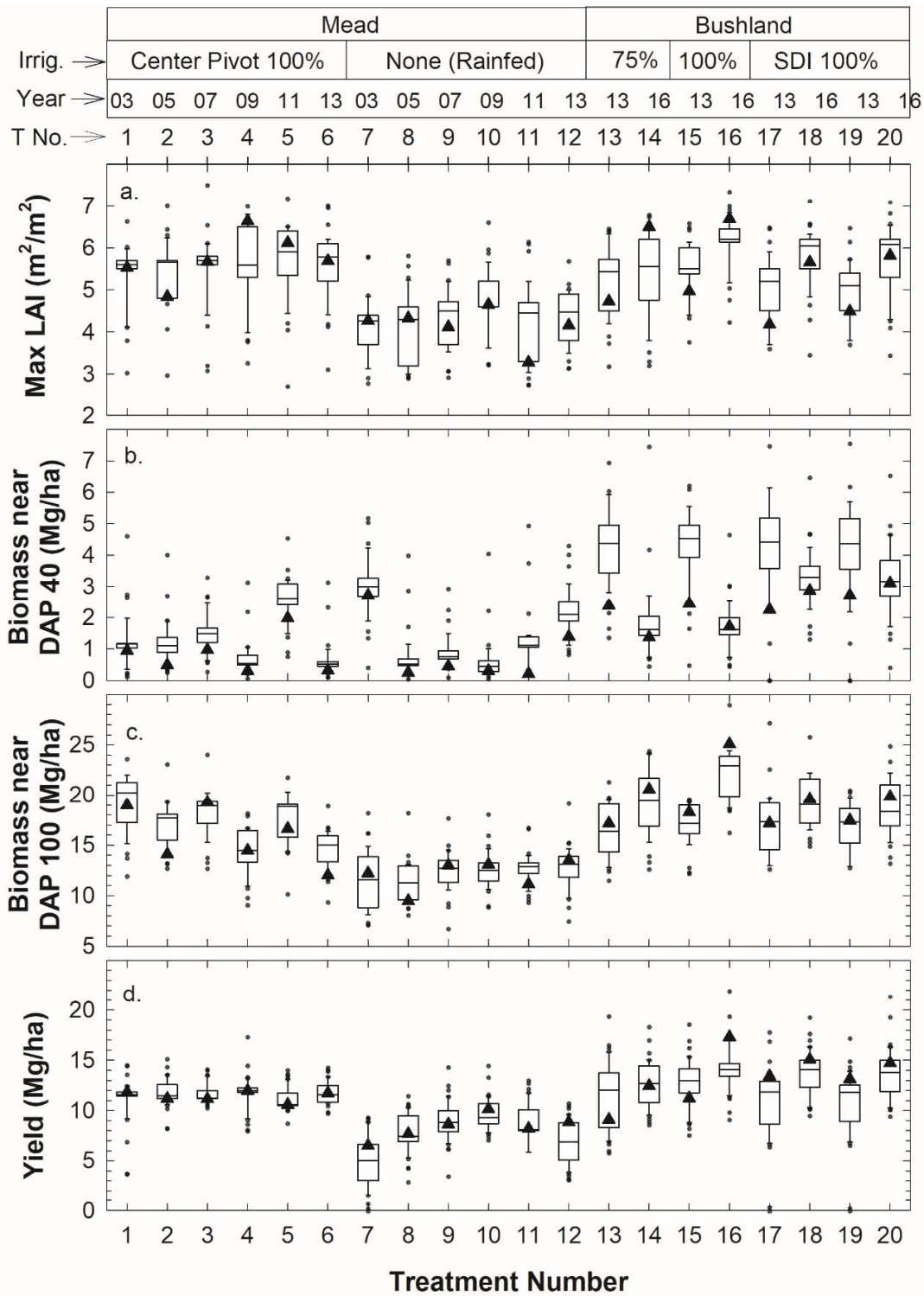
Figure 14. Box plots for Phase 4 of (a) maximum leaf area index, (b) biomass at about 40 days after planting (DAP), (c) biomass at about 100 DAP, and (d) final grain yield for all 20 treatment-years. Also shown are the corresponding observations (triangles).

55

910

911 coefficients and then alfalfa coefficients (tall, 50 cm) (Fig. 10e), which might be somewhat

912 biased because they were not independently calibrated. There was no significant difference in the

913 ability to simulate grain yield between the two methods for simulating soil E.

914

915 There was little difference in grain yield simulation ability between the two flavors of the STICS

916 model or of the MAIZSIM model (Fig. 11e). However, XNSM tended to be better than XNGM,

917 although XNGM uses a "more physiological" approach to simulate growth based on the

918 principle of functional balance, in contrast to XNSM, in which a more or less predetermined

919 scheme is used for partitioning of photosynthates.

920

921 *3.7 K-means clusters*

922 In Figs. 15a and 15c the nRMSE of simulated grain yields for 40 of the models (plus their

923 medians) and of simulated biomass accumulation for 39 of the models (plus their medians),

924 respectively, are compared against the nRMSE of the simulated cumulative ETs for the -10 to

925 +20 DAP time period (which was mostly Es for these mostly bare soil conditions). These graphs

926 show that for many of the models the relative errors for simulating ETs tended to be larger than

927 those for biomass and grain yield, which is consistent with the survey of Seidel et al. (2018) who

928 found that few modelers calibrate the ET aspects of their models. Further, k-means clustering

929 analyses with the number of clusters (k) specified to be four, the models were grouped into the

930 four clusters illustrated in Figs. 15a and 15c. As can be seen, the k-means program identified a

931 cluster of models that did quite well with the nRMSE for grain yields and biomass less than

932

933
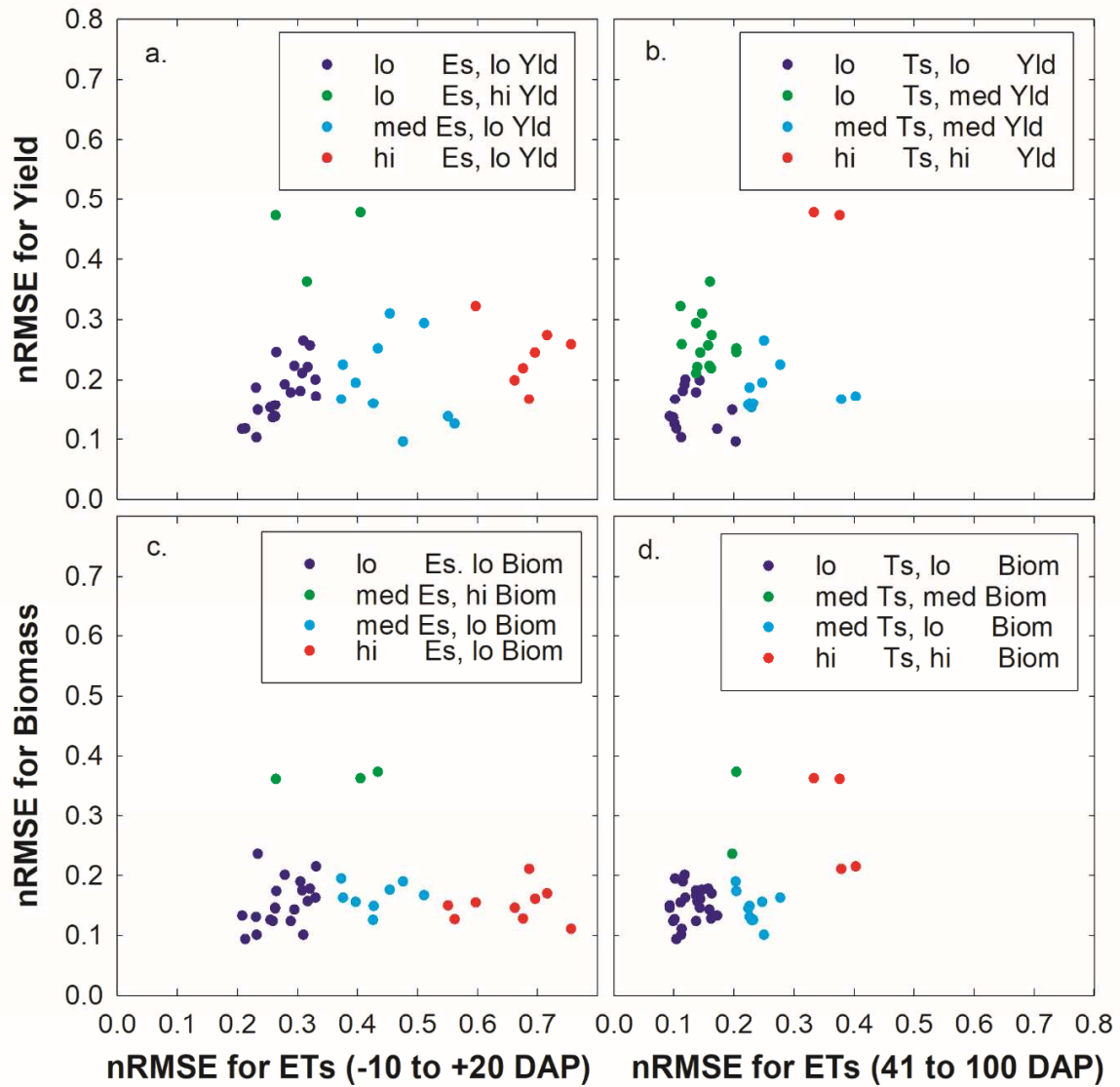
Figure 15. (a.) K-means clusters of the nRMSE for grain yields of 41 models (plus their median) for all 20 treatments versus the corresponding ETs for -10 to +20 DAP (mostly Ea). (b.) Same as (a) but for the 41 to 100 DAP periods (mostly Ta). (c.) K-means clusters of the nRMSE for biomass accumulation of 40 models (plus their median) from 41 to 100 DAP versus the corresponding ETs for from -10 to +20 DAP (mostly Ea). (d.) Same as (c) but for the 41 to 100 DAP periods (mostly Ts). (Note: one of the 41 models did not simulate grain yield and two did not simulate biomass.)

942 about 0.25 and that for ETs less than 0.35. One of the other clusters did poorly at simulating

943 grain yield and biomass, and the other two clusters did progressively worse at simulating ETs.

944 Figs. 15a and 15c suggest that a model's ability to simulate ETs well early in the growing season

945 from -10 to +20 DAP can carry on through the seasons to help simulate biomass and grain yields

946 well too.

947

948 Similarly, Figs. 15b and 15d illustrate the nRMSEs for grain yield and biomass against the

949 nRMSE for the cumulative ETs from 41 to 100 DAP, when there were mostly full crop canopies.

950 Comparing Fig. 15b with 15a and comparing Fig. 15d with 15c, it is apparent that the models

951 were better at simulating the cumulative ETs for the full canopies than they were for bare soil at

952 the beginning of the growing seasons. Again, k-means cluster analyses identified clusters of

953 models that did quite well at simulating grain yields, biomass, and full canopy ETs quite well

954 with the nRMSE of grain yield, biomass, and ETs all less than about 0.2. It is not surprising that

955 there is such a cluster of models that can simulate ETs well during midseason which aids them to

956 also simulate biomass and grain yields well.

957

958 Table 2 lists the models included in the best-performing clusters in Fig. 15. There is overlap

959 among the four categories, but CS, AMSW, ECOS, XNSM, and AHC all excelled enough to

960 appear in all four. Similarly performing well enough to appear in all four categories are three

961 flavors from the DSSAT family: DIFR, DCFR, and DIGR. Not surprisingly, the ensemble

962 median did very well, being first or second in all the categories, consistent with previous inter-

963 comparisons, e.g., Asseng et al. (2015). Among these eight models, CS, XNSM, AHC, DIFR,

964 DCFR, and DIGR all use FAO-56 (Allen et al., 1998) to compute ETp (Table S1). ETo was used

965   as ETp for DIFR and DCFR, whereas DIGR used crop coefficients to adjust ETo to ETp and

966   then various simulated or calculated crop or energy extinction coefficients were used to obtain

967   ETs. AMSW simulates T using the transpiration efficiency approach and E using Ritchie's

968   (Ritchie, 1972) two-stage method (Probert et al., 1998; Keating et al., 2003). ECOS simulates

969   ETs from net radiation that is partitioned into latent, sensible, and soil heat fluxes with energy

970   balances on the canopy and soil surfaces approach (Grant et al., 2007; Grant and Flanagan,

971   2007). DIGR uses the ASCE Standardized "Short Crop" (12-cm grass) ETo (Allen et al., 2005),

972   which is a successor to FAO-56, with maize crop coefficients computed from simulated LAI to

973   adjust ETo to ETp. Thus, six of these models have similar core approaches for simulating ETs

974   but differ in other ways such as partitioning to leaf area or soil moisture movement, etc. AMSW

975   and ECOS are both unique in their own ways within this elite group. The three DSSAT models

976   all use the Ritchie-two-stage method for soil water evaporation rather than the Sulieman method,

977   highlighting the need for E methods with improved upward movement of soil water and more

978   accurate E loss in the incomplete canopy phase.

979

980   However, something all eight models have in common is that they all have been widely used for

981   a long time under a wide range of conditions. This includes the lesser-known XNSM because it

982   is a hybrid model with elements from both the CERES model (Jones and Kiniry, 1986) and the

983   SUCROS model family (van Laar, 1992; Wang and Engel, 2000). AHC is also included because

984   it was developed based on a coupling of the significantly modified SWAP model (van Dam et

985   al., 1997) and the EPIC crop growth model (Williams et al., 1989). Thus, there has been time for

986   several generations of modelers to improve these models so that they perform well over a wide

987   range of climatic and soil conditions.

988

989

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

995

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

996

997

4.   Conclusions with discussion

998

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

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

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

4.4   Although the ensemble median was not among the best estimates of soil moisture (Supplementary), it was at the top or close to the top for all other categories. That the

1020   ensemble median generally outperforms any individual model is consistent with previous

1021   intercomparisons, e.g., Asseng et al. (2015).

1022 4.5 Within the DSSAT family, the older Ritchie (1972) approach for simulating soil E was

1023   markedly better than the newer Suleiman and Ritchie (2003, 2004) approach, which

1024   appeared to overestimate upward movement of soil moisture.

1025 4.6 Further, within the DSSAT family, the FAO-56 (Allen et al., 1998) method for calculating

1026   potential ETp was best for simulating ETs from 40 to 100 DAP (mostly canopy T) and

1027   worse for -10 to +20 DAP (mostly soil E). The Priestly and Taylor (1972) method was best

1028   for soil E and worse for canopy T. The ASCE Standardized Equation approach with short or

1029   tall crop coefficients (Allen et al., 2005) was intermediate for canopy T and worst for soil E,

1030   although this result might be somewhat biased because they were not independently

1031   calibrated.

1032 4.7 DSSAT CSM-IXIM tended to be slightly better than DSSAT CSM-CERES for simulating

1033   canopy T, probably because IXIM simulated leaf area progression better.

1034 4.8 Both STCK (which considers one surface to compute ETp) and STSW (which considers

1035   both soil and canopy surfaces to compute ETp) were among the best models to simulate ETs

1036   at the beginning of the seasons, with slightly better results for STSW. During the mid-

1037   season periods, STCK globally performed better than STSW, but both performed poorly

1038   with SDI in 2013, which distorted results.

1039 4.9 XNSM and XNGM appeared to do equally well at simulating both soil E and canopy T,

1040   with XNGM following the more detailed Farquhar modeling approach in calculating

1041   photosynthesis and leafT, but greatly simplifying vertical root distribution. However,

1042   XNSM did better than XNGM at simulating grain yield, possibly due to its simpler but more

1043   robust approach in simulating assimilate distribution among plant organs.

1044   4.9 MZD and MZH both have hourly time steps, yet MZD which uses daily weather data did

1045   slightly better than MZH which uses hourly weather data at simulating soil E, but there was

1046   no significant difference between them at simulating canopy T. This is somewhat surprising,

1047   but nevertheless shows that simulated diurnal patterns of hourly weather can be as accurate

1048   as using the actual hourly observations for input to crop growth models with hourly time

1049   steps.

1050

1066

6   Supplementary information

6.1 Word file with Table S1, which lists the ET simulation characteristics of the models plus

several figures showing an intercomparison among the models in their ability to simulate soil

moisture.

6.2 Excel file with statistics and graphs showing the Phase 4 performance of the models in their

ability to simulate the daily ET observations for the irrigated and rainfed fields in Mead in

2003 and for the 100% and 75% MESA irrigated fields in Bushland in 2013. Also included

1074   are the statistics and graphs showing the models' ability to simulate cumulative ET from -10

1075   to +20 DAP and from 41 to 100 DAP and agronomic parameters (maximum LAI, biomass at

1076   about 100 DAP, and grain yields) for all 20 treatments.

1077

References

1078

1079   Allen, R. G., Pereira, L. S., Raes, D., Smith, M., 1998. *Crop Evapotranspiration: Guidelines for*
1080   *Computing Crop Water Requirements*, FAO Irrigation and Drainage Paper 56. Food and
1081   Agriculture Organization of the United Nations, Rome, Italy.

1082

1083   Allen, R.G., Walter, I.A., Elliott, R., Howell, T., Itenfisu, D., Jensen, M., Snyder, R.L., 2005.
1084   The ASCE Standardized Reference Evapotranspiration Equation. American Society of Civil
1085   Engineers, Reston, Virginia. 195 pp.

1086

1087   Asseng, S., Ewert, F., Rosenzweig, C., Jones, J.W., Hatfield, J.L, Ruane, A.C., Boote, K.J.,
1088   Thorburn, P.J., Rötter, R.P., Cammarano, D., Brisson, N., Basso, B., Martre, P., Aggarwal,
1089   P.K., Angulo, C., Bertuzzi, P., Biernath, C., Challinor, A.J., Doltra, J., Gayler, S., Goldberg,
1090   R., Grant, R., Heng, L., Hooker, L., Hunt, L.A., Ingwersen, J., Izaurralde, R.C., Kersebaum,
1091   K.C., Müller, C., Naresh Kumar, S., Nendel, C., O'Leary, G., Olesen, J.E., Osborne, T.M.,
1092   Palosuo, T., Priesack, E., Ripoche, D., Semenov, M.A., I. Shcherbak, I., Steduto, P., Stöckle,
1093   C., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Travasso, M., Waha, M.K., Wallach, D.,
1094   White, J.W., Williams, J.R., Wolf, J., 2013. Uncertainties in simulating wheat yields under
1095   climate change. Nature Clim. Change 3, 827-832.

1096

1097   Asseng, S., Ewert, F., Martre, P., Rotter, R.P., Lobell, D.B., Cammarano, D., Kimball, B.A.,
1098   Ottman, M.J., Wall, G.W., White, J.W., Reynolds, M.P., Alderman, P.D., Prasad, P.V.V.,
1099   Aggarwal, P.K., Anothai, J., Basso, B., Biernath, C., Challinor, A.J., De Sanctis, G., Doltra, J.,
1100   Fereres, E., Garcia-Vila, M., Gayler, S., Hoogenboom, G., Hunt, L.A., Izaurralde, R.C., Jabloun,
1101   M., Jones, C.D., Kersebaum, K.C., Koehler, A.K., Muller, C., Naresh Kumar, S., Nendel, C.,
1102   O'Leary, G., Olesen, J.E., Palosuo, T., Priesack, E., Eyshi Rezaei, E., Ruane, A.C., Semenov,
1103   M.A., Shcherbak, I., Stockle, C., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Thorburn, P.J.,
1104   Waha, K., Wang, E., Wallach, D., Wolf, J., Zhao, Z., Zhu, Y., 2015. Rising temperatures reduce
1105   global wheat production. Nature Clim. Change 5, 143-147.

1106

1107   Basso, B., Ritchie, J.T., 2015. Simulating crop growth and biogeochemical fluxes in response to
1108   land management using the SALUS model, in: Hamilton, S.K., Doll, J.E., G. P. Robertson, G.P.
1109   (Eds.), The Ecology of Agricultural Landscapes: Long-Term Research on the Path to
1110   Sustainability. Oxford University Press, New York, New York, USA. pp. 252-274.

1111

1112   Bassu, S., Brisson, N., Durand, J-L., Boote, K., Lizaso, J., Jones, J.W., Rosenzweig, C., Ruane,
1113   A.C., Adam, M., Baron, C., Basso, B., Biernath, C., Boogaard, H., Conijn, S., Corbeels, M.,
1114   Deryng, D., DeSanctis, G., Gayler, S., Grassini, P., Hatfield, J., Hoek, S., Izaurralde, C.,
1115   Jongschaap, R., Kemanian, A.R., Kersebaum, K.C., Kim, S.-H., Kumar, N.S., Makowski, D.,
1116   Mueller, C., Nendel, C., Priesack, E., Pravia, M.V., Sau, F., Shcherbak, I., Tao, F., Teixeira, E.,
1117   Timlin, D., K. Waha, K., 2014. How do various maize crop models vary in their responses to
1118   climate change factors? Global Change Biol. 20, 2301-2320. doi: 10.1111/gcb.12520.

1119

1120 Best, M.J., Pryor, M., Clark, D.B., Rooney, G.G., Essery, Ménard, C.B., Edwards, J.M., Hendry,
1121 M.A., Porson, A., Gedney, A.N., Mercado, L.M., Sitch, S., Blyth, E., Boucher, O., Cox, P.M.,
1122 Grimmond C.S.B., 2011. The Joint UK Land Environment Simulator (JULES), model
1123 description Part 1: Energy and water fluxes. Geoscientific Model Development, Vol. 4, No. 3.
1124 (01 September 2011), pp. 677-699, doi:10.5194/gmd-4-677-2011

1125

1126 Brisson, N., Perrier, A., 1991. A semi-empirical model of bare soil evaporation for crop
1127 simulation models. Water Resour. Res. 27, 719-727.

1128

1129 Brisson, N., Itier, B., L'Hotel, J.C., Lorendeau J.Y., 1998. Parameterisation of the Shuttleworth-
1130 Wallace model to estimate daily maximum transpiration for use in crop models. Ecol. Model.,
1131 107, 159-169.

1132

1133 Brisson, N., Gary, C., Justes, E., Roche, R., Mary, B., Ripoche, D., Zimmer, D., Sierra, J.,
1134 Bertuzzi, P., Burger, P., Bussière, F., Cabidoche, Y.M., Cellier, P., Debaeke, P., Gaudillère, J.P.,
1135 Hènault, C., Maraux, F., Sequin, B., Sinoquet, H., 2003. An overview of crop model STICS. Eur.
1136 J. of Agron., 18, 309-332.

1137

1138 Cammarano, D., Rötter, R.P., Asseng, S., Ewert, F., Wallach, W., Martre, P., Hatfield, J.L.,
1139 Jones, J.W., Rosenzweig, C., Ruane, A.C., Boote, K.J., Thorburn, P.J., Kersebaum, K.C.,
1140 Aggarwal, P.K., Angulo, C., Basso, B., Bertuzzi, P., Biernath, C., Brisson, N., Challinor, A.J.,
1141 Doltra, J., Gayler, S., Goldberg, R., Heng, L., Hooker, J.E., Hunt, L.A., Ingwersen, J.,
1142 Izaurraldez, R.C., Müller, C., Kumar, S.N., Nendel, C., O'Leary, G., Olesen, J.E., Osborne,
1143 T.M., Priesack, E., Ripoche, D., Steduto, P., Stöckle, C.O., Stratonovitch, P., Streck, T., Supit, I.,
1144 Tao, F., Travasso, M., Waha, K., White, J.W., Wolf J., 2016. Uncertainty of wheat water use:
1145 Simulated patterns and sensitivity to temperature and CO_2. Field Crops Res., 198, 80–92. doi:
1146 10.1016/j.fcr.2016.08.015

1147

1148 Constantin, J., Willaume, M., Murgue, C., Lacroix, B., Therond, O., 2015. The soil-crop models
1149 STICS and AqYield predict yield and soil water content for irrigated crops equally well with
1150 limited data. Agric. For. Meteorol., 206, 55-68.

1151

1152 DeJonge, K. C., Thorp, K. R., 2017. Standardized reference evapotranspiration and dual crop
1153 coefficient approach in the DSSAT Cropping System Model. Trans. ASABE, 60(6), 1965-1981.

1154

1155 Doorenbos, J., Pruitt. W.O., 1985. Guidelines for predicting crop water requirements. FAO Irrig.
1156 and Drain. Paper 24. FAO, Rome.

1157

1158 Durand, J.L., Delusca, K., Boote, K.J., Lizaso, J., Manderscheid, R., Weigel, H.J., Ruane, A.C.,
1159 Rosenzweig, C., Ahuja, L., Anapalli, S., Basso, B., Baron, C., Bertuzzi, P., Deryng, D., Ewert,

1160   F., Gaiser, T., Gayler, S., Heinlein, F., Kersebaum, F.C., Kim, S.H., Muller, C., Nendel, C.,

1161   Olioso, A., Priesack, E., Villegas, J.R., Ripoche, D., Seidel, S.I., Srivastava, A., Tao, F., Timlin,

1162   D., Twine, T., Wang, E., Webber, H., Zhao, Z., 2018. How accurately do maize crop models

1163   simulate the interactions of atmospheric CO_2 concentration levels with limited water supply on

1164   water use and yield? Eur. J. Agron. (in press).

1165

1166   Evett, S.R., Heng, L.K., Moutonnet, P., Nguyen, M.L. (Eds.), 2008. Field estimation of soil

1167   water content: A practical guide to methods, instrumentation, and sensor technology. Vienna,

1168   Austria: International Atomic Energy Agency.

1169

1170   Evett, S.R., Kustas, W.P., Gowda, P.H., Anderson, M.C., Prueger, J.H., Howell, T.A., 2012.

1171   Overview of the Bushland evapotranspiration and agricultural remote sensing experiment 2008

1172   (BEAREX08): A field experiment evaluating methods for quantifying ET at multiple scales.

1173   Advances in Water Resources 50, 4-19. doi.org/10.1016/j.advwatres.2012.03.010

1174

1175   Evett, S.R., Marek, G.W., Colaizzi, P.D., Ruthardt, B.B., Copeland, K.S., 2018a. A subsurface

1176   drip irrigation system for weighing lysimetry. Appl. Eng. Agric., 34(1), 213-221.

1177   https://doi.org/10.13031/aea.12597

1178

1179   Evett, S.R., Marek, G.W., Copeland, K S., Colaizzi, P D., 2018b. Quality management for

1180   research weather data: USDA-ARS, Bushland, TX. Agrosystems, Geosciences, & Environment

1181   1(1). https://doi.org/10.2134/age2018.09.0036

1182

1183   Evett, S.R., Marek, G.W., Colaizzi, P.D., Brauer, D.K., O'Shaughnessy, S.A., 2019. Corn and

1184   sorghum ET, E, yield, and CWP as affected by irrigation application method: SDI versus mid-

1185   elevation spray irrigation. Transactions of the ASABE 62(5), 1377-1393

1186   doi.org/10.13031/trans.13314.

1187

1188   Evett, S.R., Marek, G.W., Colaizzi, P.D., Brauer, D.K., Howell, T.A., 2020. Are crop

1189   coefficients for SDI different from those for sprinkler irrigation application? Transactions of the

1190   ASAB, 63(5), 1233-1242. doi.org/10.13031/trans.13920

1191

1192   Evett, S.R., Copeland, K.S., Ruthardt, B.B., Marek, G.W., Colaizzi, P.D., Howell, T.A., Sr.,

1193   Brauer, D.K., 2022. The Bushland, Texas Maize for Grain Datasets. Ag Data

1194   Commons. https://doi.org/10.15482/USDA.ADC/1526317

1195

1196   Fleisher, D.H., Condori, B., Quiroz, R., Alva, A., Asseng, S., Barreda, C., Bindi, M., Boote, K.J.,

1197   Ferrise, R., Franke, A.C., Govindakrishnan, P.M., Harahagazwe, D., Hoogenboom, G., Naresh

1198   Kumar, S., Merante, P., Nendel, C., Olesen, J.E., Parker, P.S., Raes, D., Raymundo, R., Ruane,

1199   A.C., Stockle, C., Supit, I., Vanuytrecht, E., Wolf, J., Woli, P., 2017. A potato model

1200 intercomparison across varying climates and productivity levels. Global Change Biol. 23, 1258-
1201 1281.

1203 Gauch, H.G., Hwang, J.T.G., Fick, G.W., 2003. Model evaluation by comparison of model-based
1204 predictions and measured values. Agron. J. 95, 1442-1446.

1206 Goudriaan, J., 1977. Crop micrometerology: A simulation study. Simulation monographs.
1207 PUDOC, Wageningen, the Netherlands.

1209 Grant R.F., Arkebauer T.J., Dobermann A., Hubbard K.G., Schimelfenig T.T., Suyker A.E.,
1210 Verma S.B., Walters, D.T., 2007. Net biome productivity of irrigated and rainfed maize –
1211 soybean rotations: modelling vs. measurements. Agronomy Journal 99, 1404-1423.

1213 Grant, R. F., Flanagan, L.B., 2007. Modeling stomatal and nonstomatal effects of water deficits
1214 on CO_2 fixation in a semiarid grassland, J. Geophys. Res. 112, G03011,
1215 doi:10.1029/2006JG000302.

1217 Hasegawa, T., Lai, T., Yin, X., Zhu, Y., Boote, K., Baker, J., Bregaglio, S., Buis, S.,
1218 Confalonieri, R., Fugice, J., Fumoto, T., Gaydon, D., Naresh Kumar, S., Lafarge, T., Marcaida,
1219 M., Masutomi, Y., Nakagawa, H., Oriol, P., Ruget, F., Singh, U., Tang, L., Tao, F., Wakatsuki,
1220 H., Wallach, D., Wang, Y., Wilson, L.T., Yang, L., Yang, Y., Yoshida, H., Zhang, Z., Zhu, J.,
1221 2017. Causes of variation among rice models in yield response to CO_2 examined with Free-Air
1222 CO_2 Enrichment and growth chamber experiments. Scientific Reports | 7: 14858|
1223 DOI:10.1038/s41598-017-13582-y

1225 Hidy, D., Barcza, Z., Marjanović, H., Ostrogović Sever, M. Z., Dobor, L., Gelybó, G., Fodor, N.,
1226 Pintér, K., Churkina, G., Running, S., Thornton, P., Bellocchi, G., Haszpra, L., Horváth, F.,
1227 Suyker, A., Nagy, Z., 2016. Terrestrial Ecosystem Process Model Biome-BGCMuSo v4.0:
1228 Summary of improvements and new modeling possibilities. Geoscientific Model Development,
1229 9, 4405-4437. doi:10.5194/gmd-9-4405-2016

1231 Hoogenboom, G., Porter, C.H., Boote, K.J., Shelia, V., Wilkens, P.W., Singh, U., White, J.W.,
1232 Asseng, S., Lizaso, J.I., Moreno, L.P., Pavan, W., Ogoshi, R., Hunt, L.A., Tsuji, G.Y., Jones,
1233 J.W., 2019a. The DSSAT crop modeling ecosystem. In Boote, K. (ed.), Advances in Crop
1234 Modelling for a Sustainable Agriculture, Burleigh Dodds Science Publishing, Cambridge, UK,
1235 ISBN: 978 1 78676 240 5; www.bdspublishing.com)

1237 Hoogenboom, G., Porter, C.H., Shelia, V., Boote, K.J., Singh, U., White, J.W., Hunt, L.A.,
1238 Ogoshi, R., Lizaso, J.I., Koo, J., Asseng, S., Singles, A., Moreno, L.P., Jones, J.W., 2019b.
1239 Decision Support System for Agrotechnology Transfer (DSSAT) Version 4.7.
1240 DSSAT Foundation, Prosser, Washington. http://dssat.net.

1241

1242 Jones, C.A., Kiniry, J.R., 1986. CERES-Maize: A Simulation Model of Maize Growth and
1243 Development. Texas A&M University Press, College Station, Texas. 194 pp.

1244

1245 Keating, B.A., Carberry, P.S., Hammer, G.L., Probert, M.E., Robertson, M.J., Holzworth, D.,
1246 Huth, N.I., Hargreaves, J.N.G., Meinke, H., Hochman, Z., McLean, G., Verburg, K., Snow, V.,
1247 Dimes, J.P., Silburn, M., Wang, E., Brown, S., Bristow, K.L., Asseng, S., Chapman, S.,
1248 McCown, R.L., Freebairn, D.M., Smith, C.J., 2003. An overview of APSIM, a model designed
1249 for farming system simulation. Europ. J. Agron., 18, 267-288.

1250

1251 Kimball, B., Boote, K., Hatfield, J.L., Ahuja, L.R., Stockle, C., Archontoulis, S., Caron, C.,
1252 Basso, B., Bertuzzi, P., Constantin, J., Deryng, D., Dumont, B., Durand, J., Ewert, F., Gaiser, T.,
1253 Gayler, S., Hoffmann, M.P., Jiang, Q., Kim, S., Lizaso, J., Moulin, S., Nednel, C., Parker, P.,
1254 Palosuo, T., Priesack, E., Qi, Z., Srivastava, A., Tommaso, S., Tau, F., Thorp, K.R., Timlin, D.J.,
1255 Twine, T.E., Webber, H., Willaume, M., Williams, K., 2019. Simulation of maize
1256 evapotranspiration: An inter-comparison among 29 maize models. Agricultural and Forest
1257 Meteorology. 271:264-284.

1258

1259 Li, T., Hasegawa, T., Yin, X., Zhu, Y., Boote, K., Adam, M., Bregaglio, S., Buis, S.,
1260 Confalonieri, R., Fumoto, T., Gaydon, D., Marcaida III, M., Nakagawa, H., Oriol,P., Ruane,
1261 A.C., Ruget, F., Singh, B., Singh, U., Tang, L., Tao, F., Wilkens, P., Yoshida, H., Zhang Z.,
1262 Bouman, B., 2015. Uncertainties in predicting rice yield by current crop models under a wide
1263 range of climatic conditions. Global Change Biol. 21, 1328–1341.

1264

1265 Liu, B., Asseng, S., Müller, C., Ewert, F., Elliott, J., Lobell, D.P., Matre, P., Ruane, A.C.,
1266 Wallach, D., Jones, J.W, Rosenzweig, C., Aggarwal, P.K., Alderman, P.D., Anothai, J., Basso,
1267 B., Biernath, C., Cammarano, C.D., Challinor, A., Deryng, D., De Sanctis, G., Doltra, J., Fereres,
1268 E., Folberth, C., Garcia-Vila, M., Gayler, S., Hoogenboom, G., Hunt, L.A., Izaurralde, R.C.,
1269 Jabloun, M., Jones, C.D., Kersebaum, K.C., Kimball, B.A., A.-K. Koehler, A.-K., Kumar, S.N.,
1270 Nendel, C., O'Leary, G.J., Olesen, J.E., Ottman, M.J., Palosuo, T., Prasad, P.V.V., Priesack, E.,
1271 Pugh, T.A.M., Reynolds, M., Rezaei, E.E., Rötter, R.P., Schmid, E., Semenov, M.A., Shcherbak,
1272 Stehfest, I.E., Stöckle, C.O., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Thorburn, P., Waha,
1273 K., Wall, G.W., Wang, E., White, J.W., Wolf, J., Zhao, Z., Zhu, Y., 2016. Similar estimates of
1274 temperature impacts on global wheat yield by three independent methods. Nature Climate
1275 Change 6, 1130-1138; DOI: 10.10.38/NCLIMATE3115

1276

1277 Jones, J.W., Hoogenboom, G., Porter, C.H., Boote, K.J., Batchelor, W.D., Hunt, L.A., Wilkens,
1278 P.W., Singh, U., Gijsman, A.J., Ritchie, J.T., 2003. The DSSAT cropping system model.
1279 European Journal of Agronomy 18(3–4), 235–65. doi:10.1016/S1161-0301(02)00107-7.
1280

1281 Lopez-Cedron, F. X, Boote, K.J., Pineiro, J., Sau, F., 2008. Improving the CERES-Maize model
1282 ability to simulate water deficit effects on maize production and yield components. Agron. J.
1283 100, 296-307.

1284

1285 Maiorano, A., P. Martre, P., Asseng, S., Ewert, F., Müller, C., Rötter, R.P., Ruane, A.C.,
1286 Seminov, M.A., Wallach, D., Wang, E., Alderman, P.D., Kassie, B.T., Biernath, C., Basso, B.,
1287 Cammarano, D., Challinor, A.J., Doltra, J., Dumont, B., Rezaei, E.E., Gayler, S., Kersebaum,
1288 K.C., Kimball, B.A., Koehler, A.-K., Liu, B., O'Leary, G.J., Olesen, J.E., Ottman, M.J.,
1289 Priesach, E., Reynolds, M., Stratonovitch, P., Streck, T., Thorburn, P.J., Waha, K., Wall, G.W.,
1290 White, J.W., Zhao, Z., Zhu, Y., 2017. Crop model improvement reduces the uncertainty to
1291 temperature of multi-model ensembles. Field Crops Res., 202, 5-20.

1292

1293 Masia, S., Trabucco, A., Spano, D., Snyder, R.L., Sušnik, J., Marras, S., 2021. A modelling
1294 platform for climate change impact on local and regional crop water requirements. Agricultural
1295 Water Management 255, 107005. DOI: 10.1016/j.agwat.2021.107005

1296

1297 Mancosu, N., Spano, D., Orang, M., Sarreshteh, S., Snyder R.L., 2016. SIMETAW#—a model
1298 for agricultural water demand planning. Water Resources Management: 30, 541-557.
1299 DOI:10.1007/s11269-015-1176-7.

1300

1301 Marek, T.H., Schneider, A.D., Howell, T.A., Ebeling, L.L., 1988. Design and construction of
1302 large weighing monolithic lysimeters. Trans. ASAE, 31(2), 477-484.
1303 doi.org/10.13031/2013.30734

1304

1305 Marek, G.W., Evett, S.R., Gowda, P.H., Howell, T.A., Copeland, K.S., Baumhardt, R.L., 2014.
1306 Post-processing techniques for reducing errors in weighing lysimeter evapotranspiration (ET)
1307 datasets. Trans. ASABE 57(2), 499-515. https://dx.doi.org/10.13031/trans.57.10433

1308

1309 Monteith, J.L., 1965. Evaporation and environment. 19th Symposia of the Society for
1310 Experimental Biology, University Press, Cambridge, 19, 205-234.

1311

1312 Moore, C., Berardi, D., Blanc-Betes, E., DeLucia, E.H., Dracup, E.C., Egenriether, S., Gomez-
1313 Casanovas, S.N., Hartman, M.D., Hudiburg, T., Kantola, I., Masters, M.D., Parton, W.J., van
1314 Allen, R., von Haden, A.C., Wang, W.H., and Bernacchi, C.J., 2020. The carbon and nitrogen
1315 cycle impacts of reverting perennial bioenergy switchgrass to an annual maize crop rotation.
1316 GCB Bioenergy, 12:941-954. (http://dx.doi.org/10.1111/gcbb.12743).

1317

1318

1319 Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M,
1320 Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher,

1321   M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: Machine learning in Python. Journal of
1322   Machine Learning Research 12, 2825-2830.

1323

1324   Perego, A., Giussani, A., Sanna, M., Fumagalli, M., Carozzi, M., Alfieri, L., Brenna,
1325   S., Acutis, M., 2013. The ARMOSA simulation crop model: overall features,
1326   calibration and validation results. Ital. J. Agrometeorology 3, 23–38.

1327

1328   Penman, H.L., 1948. Natural evaporation from open water, bare soil, and grass. Proc. Royal Soc.
1329   London, 194, 120-145.

1330

1331   Priesack, E., Gayler, S., Hartmann, H.P., 2006. The impact of crop growth sub-model choice on
1332   simulated water and nitrogen balances. Nutrient Cycling in Agroecosystems 75, 1-13. DOI
1333   10.1007/s10705-006-9006-1.

1334

1335   Priestley, C.H.B., Taylor, R.J., 1972. On the assessment of surface heat flux and evaporation
1336   using large-scale parameters. Monthly Weather Rev. 100, 81-92.

1337

1338   Probert, M.E.E., Dimes, J.P.P., Keating, B.A.A., Dalal, R.C.C., Strong, W.M.M., 1998.
1339   APSIM's water and nitrogen modules and simulation of the dynamics of water and nitrogen in
1340   fallow systems. Agric. Syst. 56, 1–28. doi:10.1016/S0308-521X(97)00028-0.

1341

1342   Ritchie, J.T., 1972. Model for predicting evaporation from a row crop with incomplete cover.
1343   Water Resour. Res. 8, 1204-1213.

1344

1345   Sau, F., Boote, K.J., Bostick, W.M., Jones, J.W., Minguez, M.I., 2004. Testing and improving
1346   evapotranspiration and soil water balance of the DSSAT crop models. Agron. J. 96, 1243-1257.

1347

1348   Seidel, S.J., Palosuo, T., Thorburn, P., Wallach, D., 2018. Towards improved calibration of models –
1349   Where are we now and where should we go? European J. Agron. 94, 25-35.
1350   doi.org/10.1016/j.eja.2018.01.006

1351

1352   Shelia, V., Šimůnek, J., K.J. Boote, K.J., Hoogenboom. G., 2018. Coupling DSSAT and
1353   HYDRUS-1D for simulations of soil water dynamics in the soil-plant-atmosphere system.
1354   Journal of Hydrology and Hydromechanics 66(2), 232-245.

1355

1356   Shuttleworth, W.J., Wallace, J.S., 1985. Evaporation from sparse crops - an energy combination
1357   theory. Quart. J. Roy. Meteorol. Soc. 111, 839-855.

1358

1359   Šimůnek, J., Huang, K., van Genuchten, M., 1998. The HYDRUS code for simulating the one-
1360   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.

Šimůnek, J., van Genuchten, M. T., Šejna, M. 2008. Development and Applications of the HYDRUS and STANMOD Software Packages and Related Codes. Vadose Zone Journal, 7(2), 587. https://doi.org/10.2136/vzj2007.0077

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

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

Suleiman, A.A., Ritchie, J.T., 2003. Modeling soil water redistribution during second-stage evaporation. Soil Sci. Soc. Amer. J., 67, 377-386.

Suleiman, A.A., Ritchie, J.T., 2004. Modifications to the DSSAT Vertical Drainage Model for more accurate soil water dynamics estimation. Soil Science, 169,745-757. DOI 10.1097/01.ss.0000148740.90616.fd

Suyker, A.E., Verma, S.B., Burba, G.G., Arkebauer, T.J., Walters, D.T., Hubbard, K.G., 2004. Growing season carbon dioxide exchange in irrigated and rainfed maize. Agricultural and Forest Meteorology 124, 1-13.

Suyker, A.E., Verma, S.B., Burba, G.G., Arkebauer, T.J., 2005. Gross primary production and ecosystem respiration of irrigated maize and irrigated soybean during a growing season. Agricultural and Forest Meteorology 131, 180-190.

Suyker, A.E., Verma, S.B., 2008. Interannual water vapor and energy exchange in an irrigated maize-based agroecosystem. Agricultural and Forest Meteorology 148, 417-427.

Suyker, A.E., Verma, S.B., 2009. Evapotranspiration of irrigated and rainfed maize-soybean cropping systems. Agricultural and Forest Meteorology 149, 443-452.

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

van Dam, J.C., Huygen, J., Wesseling, J.G., Feddes, R.A., Kabat, P., van Walsum, P.E.V. et al., 1997. Theory of SWAP Version 2.0. Simulation of Water Flow, Solute Transport and Plant

1401 Growth in the Soil–Water–Atmosphere–Plant Environment, Department of water resources,
1402 WAU, Report 71, technical Document 45, DLO Winand Staring Centre-DLO.

1403

1404 van Laar, H.H., Goudriaan, J., and van Keulen, H. (eds), 1992. Simulation of crop growth for
1405 potential and water-limited production situations (as applied to spring wheat). Simulation
1406 Reports CABO-TT, no. 27, Wageningen, 72 pp.

1407

1408 Villalobos, F.J., Fereres. E., 1990. Evaporation measurements beneath corn, cotton, and
1409 sunflower canopies. Agron. J. 82, 1152-1159.

1410

1411 Wang, E. and Engel, T., 2000. SPASS: a generic process-oriented crop model with versatile
1412 windows interfaces. Environmental Modelling & Software 15, 179-188.

1413

1414 Wang, E., Martre, P., Ewert, F., Zhao, Z., Maiorano, A., Rötter, R.P., Kimball, B.A, Ottman,
1415 M.J., Wall, G.W, White, J.W, Reynolds, M.P., Alderman, P.D., Aggarwal, P.K., Anothai, J.,
1416 Basso, B., Biernath, Cammarano, D., Challinor, A.J., De Sanctis, G., Doltra, J., Fereres, E.,
1417 Garcia-Vila, M., Gayler, S., Hoogenboom, G., Hunt, L.A., Izaurralde, R.C., Jabloun, M., Jones,
1418 C.D., Kersebaum, K.C., Koehler, A.-K., Müller, C., Liu, L., Kumar, S.N., Nendel, C., O'Leary,
1419 G., Olesen, J.E., Palosuo, T., Priesack, E., Rezaei, E.E., Ripoche, D., Ruane, A.C., Semenov,
1420 M.A., Shcherbak, I., Stöckle, C., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Thorburn, P.,
1421 Waha, K., Wallach, D., Wang, Z., Wolf, J., Zhu, Y., Asseng, S., 2017. The uncertainty of crop
1422 yield projections is reduced by improved temperature response functions. Nature Plants 3 (1702),
1423 1-11. DOI: 10.1038/nplants.2017.102

1424

1425 Williams, J.R., Jones, C.A., Kiniry, J.R., and Spanel, D.A., 1989. The EPIC crop growth model.
1426 Transactions of the ASAE 32(2), 497-511.

1427

1428 Willmott, C.J. 1982. Some comments on the evaluation of model performance. Bulletin
1429 American Meteorological Society 63(11), 1309-1313.

1430

1431 Wolf, J., 2012. User Guide for LINTUL5: Simple Generic Model for Simulation of Crop Growth
1432 under Potential, Water Limited and Nitrogen, Phosphorus and Potassium Limited Conditions.
1433 Wageningen University.

1434

1435 Yang, Y., Kim, S-H, Timlin, D. J., Fleisher, D.H, Quebedeaux, B., Reddy, V. R., 2009.
1436 Simulating canopy evapotranspiration and photosynthesis of corn plants under different water
1437 status using a coupled MaizeSim+2DSOIL Model. Trans. ASAEB, 52(3), 1011-1024.

1438