Evaluation of Multimodel Averaging Approaches for Ensembling Evapotranspiration and Yield Simulations from Maize Models

Viveka Nand, Zhiming Qi, Liwang Ma, Matthew Helmers, Chandra Madramootoo, Ward Smith, T.Q. Zhang, Tobias Karl David Weber, Elizabeth Pattey, Ziwei Li, et al.

1. Introduction

Accurate prediction of crop yield and actual crop evapotranspiration (ETa) is essential for managing water resources and optimizing crop production in agriculturally dominated regions. These predictions are crucial for supporting decision-making and developing effective management strategies aimed at mitigating the impacts of natural disasters and climate change. Agricultural system models simulate the biophysical processes of crops under various climate conditions and management practices (Motha, 2011), and they help ensure sustainable crop production and enhance resilience against environmental challenges when they provide accurate and reliable predictions (Kim et al., 2021; Deb et al., 2022; Ishaque et al., 2023). Over the years, numerous crop models, ranging from simple to complex, have been developed to simulate these processes for different crops under various soil, weather, and management conditions (Kimball et al., 2023). However, the accuracy of these models in predicting crop yield and ETa remains uncertain due to potential issues with model structure, parameters, and input and calibration data (Bassu et al., 2014; Fang et al., 2019). For example, Bassu et al. (2014) study revealed that simulated maize yield varied from 10-12.5 Mg/ha, 8.5-12 Mg/ha, 6-8 Mg/ha, and 4.5-6 Mg/ha in Lusignan (France), Ames (USA), Rio Verde (Brazil), and Morogoro (Tanzania), respectively, based on 17 calibrated maize models. Similar variability in simulated maize yield and daily and seasonal ETa simulations were noted by Kimball et al. (2019) where 29 maize crop models were used. Therefore, it is challenging to determine in advance which model is most suitable for simulating crop yield and ETa across diverse climatic conditions (Martre et al., 2015; Kothari et al., 2022; Kimbal et al., 2023).

Studies on crop modeling have shown that using a combination of multiple crop models is more reliable and efficient than individual models (Bassu et al., 2014; Kothari et al., 2022; Kimbal et al., 2023). Multiple crop model ensembles help reduce errors by achieving an optimal balance between bias and variance. In these crop modeling studies, the estimated mean and median values are common ensemble predictors that equally weigh all models, demonstrating better simulation accuracy than single crop models. While weighted ensemble predictors have been suggested (Wallach et al., 2016), research is limited on the use of weighted MAAs in crop modeling. A few studies used Bayesian model averaging (BMA) (Neuman, 2003) in ensemble yield simulations and found better results than using the mean and median (Huang et al., 2017; Gao et al., 2021). Numerous other weighted MAAs, such as inverse rank, multiple linear regression (Kumar et al., 2015), machine learning algorithms (Zaherpour et al., 2019), and

Information Criterion Averaging (Akaike, 1974; Schwarz, 1978), are also discussed in the literature and used in hydrological and groundwater modeling studies.

Several hydrological and groundwater modeling studies have been performed to find the best MAAs for forecasting streamflow and groundwater levels (e.g., Ajami et al., 2006; Arsenault et al., 2015; Kumar et al., 2015; Jafarzadeh et al., 2021; Wan et al., 2021). Arsenault et al. (2015) compared nine MAAs using 12 hydrographs (4 models × 3 metrics) across 429 catchments and found that the MLR B method outperformed others, with no catchment requiring more than seven ensemble members to obtain better stream flow simulation. Similarly, Kumar et al. (2015) evaluated ten different MAAs methods using eight hydrological models to determine the best method for discharge estimation in the Mahanadi River basin in India and concluded that MLR C was the most suitable MAAs method, with five model ensembles providing the best discharge simulations for the study area. We hypothesize that those MAAs can also improve the simulation accuracy of the crop yield and ETa by an ensemble of agricultural systems models. However, to our knowledge, these MAAs have not yet been applied in crop modeling studies to compute ensemble yield and daily ETa estimates from simulations.

Crop yield and ETa simulation accuracy can be increased by calibrating crop model parameters using various observed data sources. These include field experimental data, such as initial water content, phenological events, soil water content, leaf area index (LAI), daily ET_a, biomass, and yield. However, these measured data sets are often not available at all sites, and the limited availability of measured data can remarkably impact the predictive capabilities of individual crop models in predicting crop yields and ET_a. In past maize modeling studies, the mean or median of yield and daily ETa simulations were satisfactory under blind (uncalibrated) and calibrated applications; however, the best MAAs approach to improving simulations is still needed. The present study evaluates the performance of seven MAAs and identifies the best MAA approach to estimate maize yield and daily ET_a for blind and calibrated model applications across several locations in the USA and Canada. The study utilized two simulation data sets: Group A) five maize models were used in this study to simulate maize yield and daily ETa compared to measured data from nine US and Canadian sites, and Group B) the simulation results from previous AgMIP study (Kimball et al., 2023) for 41 maize models used to simulate maize yield and daily ETa at Mead, Nebraska and Bushland, Texas were evaluated. AgMIP (Agricultural Model Intercomparison and Improvement Project) is a global initiative focused on improving agricultural system models to better assess the impacts of climate change, economic shifts, and social factors on agriculture (https://agmip.org/agmipcharter2/). By uniting scientists and comparing multiple models against real-world data, AgMIP enhances the accuracy of predictions for key crops such as maize, wheat, and rice, helping to inform strategies for food security and agricultural resilience.

2. **Materials and Methods**

2.1 Description of field experiment sites and experiment data

Nine maize (Zea mays L.) field experiment sites (Group A) were selected for analysis: Ames (Iowa, USA), Gilmore (Iowa, USA), Greeley (Colorado, USA), Ithaca (Nebraska, USA), Glenlea (Manitoba, Canada), Harrow (Ontario, Canada), Ottawa (Ontario, Canada), Sainte-Anne-de-Bellevue (Quebec, Canada), and Saint Emmanuel (Quebec, Canada) (Table 1 and Fig. 1). In addition, two maize field sites (Group B) previously used for AgMIP maize project ETa and yield simulations studies (Mead and Bushland) were selected, focusing on four treatments (i.e., Mead rainfed, Mead irrigated, Bushland 75% Mid Elevation Sprinkler Application (MESA) irrigation, Bushland 100% MESA irrigation). The Bushland, Mead, Ithaca, and Greeley sites were irrigated while the remaining sites were rainfed. The average growing season air temperature, rainfall, and soil types of each site are given in Table 1. The average growing season temperature varied between 10.40°C in Ithaca, USA, and 22.80°C in Bushland, USA, while seasonal precipitation ranged from 191 mm in Greeley, USA, to 592.36 mm in Ithaca, USA across the maize experiment sites.

A detailed description of available measurements of each site is given in Supplementary Information Table 1. In-situ measured daily weather data, including maximum and minimum air temperature, rainfall, wind speed, relative humidity, and solar radiation, were utilized for all sites except Sainte-Anne-de-Bellevue, where specific site weather data were not measured. Weather data for Sainte-Anne-de-Bellevue was obtained from the nearest weather station of Environment Canada. For soil-related information, measured soil profile data were used across all sites. Comprehensive crop management details, including tillage practices, cultivar details, seeding rate, seeding date, plant density, fertilizer application rate, harvesting date, biomass, and grain yield were obtained for all sites. The quantity and timing of irrigation was obtained for the irrigated sites. Phenological dates, detailing the various stages of plant development, were meticulously recorded for Ames, Bushland, Greeley, Mead, Ottawa, and Saint Emanuel. Additionally, time-series measurements of Leaf Area Index (LAI) and actual crop evapotranspiration (ETa) were obtained for Ames, Bushland, Greeley, Mead, and Ottawa. Measured layer-wise soil water content data were available for all sites except Harrow and Sainte-Anne-De-Bellevue.

2.2 Crop model setup and calibration

As mentioned in section 2.1, we utilized two types of crop yield and ETa data sets. The first set named "Group A" was comprised of simulated crop yield and ETa data from the uncalibrated (Blind Phase) and fully calibrated phases of the five maize models in this study (Table 1). The second set (Group B) included simulated daily ETa and yield data from uncalibrated and fully calibrated phases of 41 maize models for the Bushland and Mead sites. This data was sourced from the Agricultural Model Inter-comparison and Improvement Project (AgMIP; https://agmip.org/). The description of 41 Maize Models is given in Supplementary Information Table 2. A detailed explanation of the model set-up and calibration process is presented in Kimball et al. (2023).

The five best Maize crop models, as selected from the AgMIP studies were used to simulate crop yield and ETa for Group A's sites. (Supplementary Information Table 3). These include DSSAT-CERES maize with Priestly-Taylor Ritchie ET equation (DCPR), DSSAT-CERES maize with FAO56 Ritchie ET equation (DCFR), APSIM-maize with SOILWAT Archontoulis subroutine (AMW), APSIM-maize with SWIM Archontoulis subroutine (AMSA), and RZWQM2. The selection of a combination of crop models was based on an AgMIP project in which 29 maize crop models were compared (Kimball et al., 2019). Maize yield predictions were calibrated and validated using measured field data (Kimball et al.,2019). The RZWQM2 model which uses the Shuttleworth-Wallace approach to estimate potential transpiration (PT) and potential evaporation (PE) (Shuttleworth and Wallace, 1985) did not perform well in simulating ETa among the five crop models, however, it was in the top five in simulating crop yield and therefore included in this study. Models were set up utilizing site-specific measured data, encompassing layered soil texture along with corresponding physical and hydraulic properties, tillage dates, cultivar details, seeding dates, plant density, irrigation amounts, and fertilizer rates.

In the blind phase, for Group A sites, all five maize models were set up using site-specific measured input data, including soil, weather, and crop management details (such as seeding date, plant density, and fertilizer rate). The models' phenology parameters were then adjusted to align with the crop maturity dates across all sites. Subsequently, the models were run to simulate ETa and yield. During this phase, models were not calibrated with available soil moisture, ETa, and yield data. In the calibrated phase, however, all maize models were fine-tuned against the measured data to improve their ETa and crop yield simulation accuracy. Cultivar parameters in each model were initially adjusted to align anthesis, silking, and maturity dates with observed ones depending on sites and available phenological measurement dates. Subsequently, the models underwent calibration against soil water content data by adjusting saturated and lateral hydraulic conductivity for all sites except Harrow and Sainte-Anne-de-Bellevue. Following this, the models were fine-tuned for ETa by adjusting parameters related to albedo, soil resistance, and leaf stomatal resistance at sites with ETa measurements. Lastly, the models were calibrated for leaf area index (LAI) for those sites (Ames, Bushland, Mead, Ottawa, and Sainte-Anne de Bellevue) that had LAI observations and crop yield by adjusting cultivar parameters influential on crop yield. Among the field experiment sites, ETa was simulated for Greeley, Ames, and Ottawa as observed maize ETa data was available only for these sites. Crop yield was simulated for all sites.

2.3 Model Averaging Approaches (MAA)

The simulated yield and daily ETa data from all sites were ensembled using seven MAAs: Simple Model Averaging (SMA), Median, Inverse Rank (IR), Bates and Granger Averaging (BGA), and three variants of Granger Ramanathan (MLR A, MLR B and MLR C), (Supplementary Information Table 4). First, the simulated yield and daily ETa from all maize models were combined using all seven model averaging methods. We referred to it as "all maize models". Then, the simulated yield and daily ETa of one flavor model from each model family were selected and ensembled. It was named "group maize models". The simulated yield was averaged across all sites, while the simulated ETa values were averaged for three sites of Group A (Ames, Greeley, and Ottawa) and all sites of Group B. SMA, Median, IR, BGA, MLR A, MLR B, and MLR C were applied to all sites to estimate the weight of each maize model. The average yield and ETa were then determined by multiplying the weight of each maize model with its corresponding simulated yield and daily ETa for each site. The resulting yields and daily ETa obtained through multimodel average methods were subsequently compared with observed yield and daily ETa sets. Details of the multiple MAAs are given below:

a. **Simple Model Averaging (SMA):** In this approach, the weight of each model is assigned equally. Mathematically, it can be estimated as:

$$W = \frac{1}{n} \tag{1}$$

Where n is the number of ensemble models, and W is the estimated weight of each model.

b. **Median:** The median of simulated values of all ensemble models is taken to combine the forecast.

c. **Inverse Rank:** The inverse rank approach, rank the model simulation based on their performance. The first rank is assigned to model with lowest mean squared error, the model with the second lowest mean squared error is assigned the rank 2. Then weightage of each model is calculated as follows:

$$W = \frac{Rank_i^{-1}}{\sum_{i=1}^{N} Rank_i^{-1}} \tag{2}$$

d. **Bates and Granger Averaging (BGA):** The BGA method combined the forecast of ensemble models by minimizing the mean square error between simulated and observed values. It can be estimated as:

$$W = \frac{\frac{1}{RMSE^2}}{\sum_{i=1}^{N} \frac{1}{RMSE^2}} \tag{3}$$

Where RMSE is the root mean square error of the i^{th} ensemble model.

e. **Granger Ramanathan (MLR A, MLR B, and MLR C):** The MLR A approach, developed by Granger and Ramanathan in 1984, employs the ordinary least squares (OLS) method to assign weights, effectively lowering the root mean square error (RMSE) but lacking bias correction. MLR B is similar to MLR A but includes a bias correction mechanism. Conversely, MLR C uses constrained least squares, ensuring that the weights of all models sum to one. In MLR C, weights are estimated by:

$$W = (Q^T{}_{sim} Q_{sim})^{-1} Q^T{}_{sim} Q_{obs} \tag{4}$$

Where Q_{obs} is the matrix of the observed values, Q_{sim} is the matrix of simulated values, and $Q^T{}_{sim}$ is the transpose matrix of simulated values.

2.4 Performance Evaluation of the Models

The evaluation of the crop models and model averaging methods performance was assessed by statistical indicators such as relative root mean square error (RRMSE). Jamieson et al. (1991) concluded that RRMSE values below 10% are “excellent”, values from 10-20% are “good”, values from 20-30% are “satisfactory”, and values exceeding 30% are “poor”.

$$RRMSE = \frac{100}{\bar{o}}\sqrt{\frac{1}{n}\sum_{i=1}^{n}(o_i - s_i)^2} \quad (5)$$

Where n is the number of observed and simulated data points, o_i is the observed value, s_i is the model simulated value, $\bar{o}$ is the mean of observed values.

3. Results

3.1 Group A Sites Simulations

In this section, the simulated daily ETa and seasonal yield were examined using five maize crop models (DSPR, DSFR, AMW, AMSA, and RZWQM2) across nine sites in the USA and Canada, under both the blind and calibrated phases. Additionally, the MAAs' estimated daily ETa and seasonal yield results were assessed. The analysis focused on daily ETa simulations at Ames, Greeley, and Ottawa, where daily ETa measurements were available. Seasonal yield was analysed at all nine sites. For Ames, Greeley, and Ottawa, the analysis focused on the growing seasons of 2006, 2010, and 2006 for daily ETa simulations, respectively.

3.1.1 Blind phase

Crop Evapotranspiration

A wide range of daily ETa simulations was observed in the five maize models at all sites, especially in the mid and end-growth stages during the blind phase (Fig. 2). The RRMSE between measured and simulated daily ETa ranged from 47.5-63.6% at Ames, from 36.5-104.2% at Greeley, and from 34.5-75.4% at Ottawa (Fig. 3a). In 2006 at Ames, the measured average daily ETa during the growing season was 2.5 mm, while the simulated average daily ETa ranged from 2.3-2.7 mm/day. Similarly, at Greeley in 2010, the measured average daily ETa was 4.4 mm, and simulated average daily ETa values ranged from 3.6-6.9 mm/day. In Ottawa in 2006, the measured average daily ETa was 2.3 mm, while simulated values varied between 2.2-3.3 mm/day.

However, ensembling the daily ETa simulations from all five maize models using seven model averaging methods improved the accuracy of daily ETa simulations based on the RRMSE (Fig. 3a and Table 2). The performance of MLR C model averaging methods to combine daily ETa simulations was best at the Ames and Greely sites, whereas MLR A performed slightly better at the Ottawa site. Figure 2 indicates closer agreement between measured and MLR C ensembled daily ETa over the growing season at all sites.

When daily ETa simulations of group maize models were ensembled, the performance of model averaging methods decreased compared to the ensembling of all maize models (Table 4). Though MLR B and MLR C model averaging methods showed almost similar performance in combining daily ETa, MLR B ensemble daily was best at the Greeley and Ottawa sites, whereas MLR C performed best at the Ames site.

Crop Yield

Uncalibrated maize models showed unsatisfactory performance across all sites, as indicated by high RRMSE values (Fig. 4a). However, combining simulated yields from all maize models using model averaging methods remarkably improved yield simulation performance, achieving acceptable RRMSE criteria. Generally, the performance of MLR A and MLR B was similar across all sites, followed by MLR C, IR, BGA, SMA, and the Median (Fig. 4). Additionally, when yield simulations from group maize models were ensembled, no improvements were found as compared to an ensemble of all maize models (Table 3). There was a slight decrease in the performance of the model averaging method in the ensemble of group maize models.

3.1.2 Calibrated Phase

Crop Evapotranspiration

Substantial variability in the daily simulated ETa persisted at each site, despite calibrating all crop models (Fig. 5). The RRMSE values ranged from 45.2-52.4% at Ames, 36.4-53.7% at Greeley, and 34.6-71.5% at Ottawa (Fig. 3b), indicating that the RRMSE remained in the unacceptable range across all maize models and sites. At the Ames site, the average measured growing season daily ETa was 2.5 mm, while the average simulated daily ETa ranged from 2.4-3.0 mm/day across all maize models. Similarly, in Greeley, the average growing season measured daily ETa was 4.4 mm, with simulated values between 3.7-4.5 mm/day. Similar results were observed at the Ottawa site. However, when an ensemble of all maize models was taken using model averaging methods, this variability was reduced across all sites as shown by

RRMSE values in Fig 3b. A slightly improvement in ensembled daily ETa simulations was noted across all model averaging methods compared to the blind phase (Table 2). The RRMSE for the ensemble varied from 37.1-49.9% at Ames, 26.4-33.4% at Greeley, and 29.7-38.3% at Ottawa across all MAAs. The MLR C ensemble of daily ETa showed closer agreement with the measured daily ETa than other MAAs at all sites. Furthermore, the accuracy of daily ETa improved when averaging group maize models compared to averaging all maize models (Table 2). MLR C performed the best for combining daily ETa at Ames and Greeley, while MLR B was the best at the Ottawa site.

Crop Yield

When all maize models were fully calibrated, their performance improved across all sites. Comparing the simulated yields of individual maize models with the measured yields, the RRMSE was found to be less than 30% (Fig.4b), indicating that the performance of each crop model varied depending on the site, and no single model consistently outperformed others for simulating maize yield across all locations. The RRMSE between measured and simulated yield ranged from 0.44% to 28.90% across all maize models and sites.

Yield simulations improved further when an ensemble of all maize models was taken using model averaging methods, as indicated by RRMSE values in Fig. 4b. The MLR A produced ensembled yield values were very close to the observed yields at all sites. The performance of MLR B was comparable to MLR A at most sites with slight variation. In the calibrated phase, the performance of model averaging methods was slightly better than in the blind phase.

However, a minor decrease in the accuracy of yield simulations was noted when using an ensemble of group maize models with model averaging methods, indicating that the ensemble of simulated yield from group maize models did not improve the yield simulations (Table 3). Among the model averaging methods, the ensemble yields from MLR A and MLR B matched the measured yields at most sites.

3.2 Group B Sites Simulations

3.2.1 Blind Phase

Crop Evapotranspiration

The 41 maize models from the AgMIP maize ET study simulated daily ETa were in a wide range at all sites (Kimball et al., 2023). The RRMSE between the daily simulated ETa and the in-situ measured daily ETa ranged from 33% to 110% at Mead irrigated, 32% to 131% at Mead rainfed, 29% to 87% at Bushland 100% MESA, and 31.20% to 79% at Bushland 75% MESA sites across all maize models (Fig.6a). The previous analysis by Kimball et al. (2023) revealed that the median of all maize models closely matched the measured daily ETa throughout the growing season. In the present study, variability in daily ETa simulations decreased when the ensemble of all maize models was used. Even though roughly similar performance was noted for the MLR A, MLR B, and MLR C at all sites except Bushland 75% MESA, overall, MLR C-ensembled daily ETa performed better in matching the daily measured ETa over the growing season at most sites, followed by MLR A, MLR B, IR, BGA, SMA, and the Median (Table 4). The RRMSE between the ensembled daily ETa and the measured daily ETa ranged from 18.4% to 28% at Mead irrigated, 18.5% to 38.1% at Mead rainfed, 19% to 26.4% at Bushland 100% MESA, and 25.8% to 30% at Bushland 50% MESA sites in among MAAs (Table 4 and Fig.6a).

The ensembled daily ETa was also compared using SMA and MLR C with the measured daily ETa during the 2003 growing season at Mead's irrigated and rainfed sites. Fig. 7 illustrates a close match between the measured daily ETa and the MLR C ensembled daily ETa, particularly towards the end of the growing season at the Mead Irrigated site. The MLR C ensembled daily ETa followed the pattern of the measured daily ETa more closely than the SMA ensembled daily ETa. However, none of the MAAs could reproduce the peak daily measured ETa. Similarly, at the Mead rainfed site, the MLR C ensembled daily ETa closely followed the daily measured ETa for the 2003 growing season (Fig.7), whereas the SMA ensembled daily ETa showed poor agreement with the measured daily ETa, especially during the mid-and late-growing seasons. MLR C ensembled daily ETa also closely followed the pattern of daily measured ETa during the 2013 crop period at Bushland 100% MESA and 75% MESA sites. However, the MLR C and other MAAs underestimated ETa during the early and mid-crop periods. This discrepancy is attributed to the inadequacy of many crop models in accounting for varying wind speed and humidity. The models estimated ETa accurately during periods of lower ETa but considerably underestimated ETa during periods of higher ETa, characterized by high wind speeds and low relative humidity (Kimball et al., 2023).

Additionally, the results of group maize models were analyzed, where one model from each crop model family was selected. This approach marginally improved the daily ETa simulations at all sites compared to considering an ensemble of all maize models (Table 4). For instance, the RRMSE between the daily measured ETa and the ensembled daily ETa of all maize models ranged from 18.4% to 28% across all models averaging methods at the Mead irrigated site. In contrast, the RRMSE between the daily measured ETa and the ensembled ETa of group maize models ranged from 18.6% to 24.4% across all model averaging methods. Similar findings were observed at the Mead rainfed, Bushland 100% MESA, and Bushland 75% MESA sites.

Crop Yield

Large variability in simulated maize yields was noted across 41 maize models during the blind phase (Fig. 8a). An ensemble of simulated yields of all maize models reduced the deviation between measured yield and simulated maize yield at all sites. Among the seven MAAs, MLR A performed the best followed by MLR B, MLR C, IR, BGA, SMA, and median at most sites. Moreover, the performance of group maize models was examined. Overall, this approach improved the yield simulations for a few cases (Table 4). The performance of all MAAs in combining the simulated yield of group maize models was roughly similar to ensembling the maize yield of all maize models.

3.2.2 Calibrated Phase

Crop Evapotranspiration

After fully calibrating all maize models, a slight improvement in daily ETa simulations was noted in all maize models. There was still wide variability in daily ETa simulations across the 41 maize models. The RRMSE ranged from 28.5% to 75.0%, 30.3% to 90.0%, 30.0% to 68.5%, and 28.0% to 67.0% at Mead irrigated, Mead rainfed, Bushland 100% irrigation, and Bushland 75% irrigation sites, respectively (Fig. 6b). Model averaging methods reduced the variability in daily ETa simulation by ensembling daily ETa simulations of all maize models. In the calibrated phase, improvement in ensembled daily ETa simulation across MAAs was slightly higher than in the blind phase at all sites (Table 4). Though MLR A, MLR B, and MLR C MAAs showed almost similar performance to ensemble daily ETa of all maize models, MLR A outperformed others at Mead rainfed and irrigated sites and MLR C outperformed others at Bushland 75 and 100% MESA sites. For instance, the RRMSE between the MLR A ensembled daily ETa and measured daily ETa was 19.0 and 19.4% at Mead irrigated and rainfed sites, respectively (Fig.6b). Similarly, RRMSE between the MLR C ensembled daily ETa and measured daily ETa was noted for 19.30% and 19.40% at Bushland 100% MESA and 75% MESA sites, respectively The model averaging methods ensembled daily ETa were also compared with measured daily ETa over the growing season at Mead and Bushland sites. Fig. 9 shows a close match between in-situ measured daily ETa and MLR C ensembled daily ETa, particularly during the 2003 growing season at Mead rainfed, where MLR C closely followed the measured pattern.

Moreover, the ensemble of daily ETa of group maize models was compared using different model averaging methods. A slight improvement in ensembled daily ETa simulations was noted when considering group maize models (Table 4), however, the pattern of performance of MAAs to ensemble daily ETa simulations of group maize models was similar to all maize models. For example, MLR A model averaging method ensembled daily ETa was found best at Mead irrigated and rainfed sites, whereas MLR C ensembled daily ETa outperformed to others at Bushland 100 and 75% MESA sites in both cases (Table 4).

Crop Yield

Simulated yield showed remarkable improvement in most maize models after full calibration compared to the blind phase (Fig 8b). The greatest improvement in yield simulations was observed at the Mead irrigated site; however, moderate variability in yield simulations was found across all maize models at the Mead rainfed, Bushland 100% MESA, and Bushland 75% MESA sites. This variability decreased substantially when simulated yields were averaged using model-averaging methods at all sites. The MLR A performed the best at all sites, followed by MLR B, MLR C, IR, BGA, SMA, and the median. The RRMSE between simulated and measured yields ranged from 0.03-4.0% at Mead irrigated, 5.6-12.8% at Mead rainfed, 4.2-15% at Bushland 100% MESA, and 2.8-19% at Bushland 75% MESA sites across all model-averaging methods (Table 4). Additionally, the ensembling of simulated yield from group maize models showed mixed results compared to combining simulated yields from all maize models across all model-averaging methods. There was a marginal improvement in yield simulation at Mead rainfed and Bushland 75% MESA sites compared to all maize models, while there was a slight decrease noted at Mead irrigated and Bushland 100% MESA sites (Table 4).

4. Discussion

Blind vs Calibrated

Combining simulations from multiple models through various model-averaging approaches often provides more accurate simulation performance (Sandor et al., 2023). In this study, as anticipated, MAAs performed slightly better during the calibrated phase than for the blind phase for combining ETa and yield simulations of all and group maize models (Table 5, 6). In crop modeling, calibrated is a crucial process aimed at estimating unknown parameters using field observations, thereby reducing uncertainty in model simulations and making predictions more reliable (He et al., 2017). MAAs tend to perform better in the calibrated phase because the models are fine-tuned to specific datasets, which minimizes errors and variance, resulting in more accurate and stable predictions (Fletcher, 2018).

Interestingly, MAAs also performed well in the blind phase. The outcomes of the present study are comparable to those of Bassu et al. (2014) and Kimball et al. (2019), where the maize yield and ETa simulations from uncalibrated maize models in different climatic conditions sites were combined using the mean and median. However, in this study, an additional five MAAs were tested, which will be discussed in the next section. Similarly, Ajami et al. (2006) found that averaging streamflow simulations of uncalibrated multiple hydrological models using four model combination methods performed better than a calibrated single hydrological model. These studies found that multi-model combinations could enhance prediction accuracy by compensating for individual model errors to reduce variance (Bassu et al., 2014; Kimball et al., 2019; Kimball et al., 2023; Sandor et al., 2023; Couëdel et al., 2024). The multi-model combination improves the simulation accuracy by reducing the variance associated with the predictions (Bassu et al., 2014; Fletcher, 2018). The individual model might exhibit high variance due to their sensitivity to model structures and parameters. By averaging the outputs of multiple models, these variances are reduced, leading to more stable and reliable predictions. In addition, different models may make different errors when predicting. When these models are averaged, the errors can cancel each other out to some extent, resulting in a more accurate overall prediction. Nonetheless, while multi-model ensembles offer a way to learn from the errors across various studies and improve the models, some individual models might still outperform the mean and median (Kothari et al., 2022).

Best Model Averaging Method for ETa and Yield

The study assesses how well different MAAs can reduce variability and improve the accuracy of daily ETa and yield simulations at various Group A and Group B sites. Remarkably, SMA and the median approach performed better than individual calibrated maize models in 98% of the cases during the blind phase at Group A sites, with SMA usually outperforming the median. Similar results were observed in Group B sites for ETa and yield. This could be due to a trade-off in prediction errors among different models, leading to more accurate overall predictions. These findings are comparable to that of Ajami et al. (2006), Bassu et al. (2014), Arsenault et al. (2015), Sandor et al. (2023), and Couëdel et al. (2024) which showed that the mean of simulated streamflow and yield from hydrological and crop models, respectively, was better than individual calibrated models.

Further enhancement in daily ETa and maize yield simulations was noted when other model averaging methods, such as IR, BGA, MLR A, MLR B, and MLR C, were used. Overall, the improvements ranged between 3.5-6.5% for daily ETa and 3.3-9.7% in terms of RRMSE for yield simulations at Group A sites across the five MAAs compared to the median (Table 5). Similarly, improvements in daily ETa and yield simulations ranged between 3.2% and 8.7%, and 7.3% and 9.5%, respectively, at Group B sites (Table 6).). The improvement in daily ETa and yield estimations by the additional five MAAs over the median was slightly greater for daily ETa and moderately greater for yield in the blind phase compared to the calibrated phase (Table 5 and Table 6). BGA often performed better in combining daily ETa simulations than SMA and the median, though it was usually outperformed by its variant IR (Table 5, 6). This can be explained by the IR method's disregard for outliers (Aiolfi and Timmermann, 2006). For yield simulations, BGA and IR showed almost similar performance. According to Diks and Vrugt (2010), BGA did not outperform other methods (AICA, BICA, BMA, and MLR A) except SMA.

When comparing the performance of MLR A, MLR B, and MLR C, there were only marginal differences in their ability to combine daily ETa and yield simulations in 75% of cases, aligning with the study by Arsenault et al. (2015) (Table 2, 4). MLR A, MLR B, and MLR C performed considerably better than SMA and the median and slightly to moderately better than IR and BGA, depending on the site. Overall, averaging the RRMSE of all sites for all maize models and group maize models for blind and calibrated phases revealed that MLR C was best for daily ETa simulations, while MLR A was best for yield simulations (Table 5, 6). MLR C improved daily ETa estimation by an average of 6.5% and 8.7% in terms of RRMSE than the median, while MLR A enhanced maize yield estimation by 9.8% and 9.2% for Group A and Group B sites, respectively.

This is likely because of higher bias in daily ETa simulations across maize models compared to yield simulations. MLR A was better at reducing variance in yield simulations due to incorporating variance reduction. In contrast, MLR C reduces variance by giving positive higher weights to well-performing models while minimum weight to the worst-performing models even in some cases zero. Therefore, it combined the daily ETa simulations slightly better than other MAAs. For ETa, the results were contradicted by Ajami et al. (2006), Arsenault et al. (2015), and Wan et al. (2021) and comparable to Kumar et al. (2015).

Kumar et al. (2015) found that MLR C was the best method for combining simulated river discharge from eight hydrological models. For crop yield, findings were in line with (Diks and Vrugt, 2010), who reported that MLR A's results were similar to advanced MAAs such as Bayesian Model Averaging (BMA) and Mallows Model Averaging (MAAS). The advantage of using MLR A over BMA or MAAS can be notable since MLR A has straightforward solutions for determining weights. In contrast, finding the best weights for BMA and MAAS requires more complex and time-consuming methods, such as the Differential Evolution Adaptive Metropolis (DREAM) adaptive Markov chain Monte Carlo (MCMC) algorithm.

Overall, the MLR A and MLR C methods were found to outperform others for ensemble yield and ETa simulations of maize models, respectively, in both data sets. This emphasizes the importance of selecting appropriate averaging techniques. The success of these methods can be attributed to their ability to integrate multiple model outputs, leveraging the strengths and compensating for the weaknesses of individual models.

Moreover, ensemble group maize models improved the simulation accuracy of crop yield and ETa in a few cases compared to ensemble all maize models. However, the accuracy of the ensembled ETa and yield simulation of group maize models was similar to that of the ensembled ETa and yield simulation of all maize models. This finding suggests that the diversity of models in the ensemble plays a crucial role in enhancing prediction accuracy. Therefore, it is advisable to select ensemble members from different crop family models to achieve the best results, although its also true that the quality of modelers regarding the assumptions they make in parameterizing models is also of importance (Albanito et al., 2022).

Model Averaging Methods when "No Observations Data" is available

Most MAAs, such as IR, BGA, MLR A, MLR B, and MLR C, typically rely on ground measurement data to determine the weights for each model in the ensemble. This data is crucial for selecting the best models and assigning appropriate weights. However, in real-world scenarios, experimental data not be available, posing substantial challenges for model selection and weighting.

In such situations, SMA and the median method have shown promising results. SMA and the median method are straightforward approaches that average predictions from multiple models by assigning equal weights to each. This simplicity is particularly advantageous when there is no prior information about the performance of the individual models. By averaging the outputs, SMA reduces the impact of biases or errors from any single model, leading to more robust overall predictions. Both methods were effective in the current study, where they combined multiple crop model outputs to improve predictions of daily ETa and yield, even in the blind phase. This finding is consistent with previous crop modeling studies by Bassu et al. (2014), Martre et al. (2015), Kothari et al. (2022), Kimball et al. (2019, 2023), who reported that the mean and median of ETa and yield simulations from multiple crop models often outperform individual crop models.

However, the main drawback of SMA and the median method is that they do not fully leverage the strengths of the better-performing models. Because all models are weighted equally, these methods may underutilize the models that have superior predictive capabilities. Despite this limitation, SMA and the median method remain valuable tools in scenarios where observational data are lacking, providing a practical means of improving predictive accuracy by mitigating individual model weaknesses.

5. Conclusions

Averaging the results from multiple agricultural systems models has shown high accuracy in predicting crop yield and ETa. However, among those available Model Averaging Approaches (MAAs), it is not known which one performed the best. Therefore, this study aimed to evaluate the performance of seven MAAs (SMA, Median, IR, BGA, MLR A, MLR B, and MRL C) across eleven sites in North America to predict maize yield and daily ETa using two ensemble-size maize crop models (all maize models and group maize models) and two calibration approaches (Blind and Calibrated phases). The data come from two sources: simulations for Group A sites were done in this study, while simulations for Group B sites were carried out by the Maize AgMIP project team.

The following conclusions were drawn from the study:

- **Model Averaging Approaches:** All MAAs (Model Averaging Approaches) generally performed well, often surpassing individual crop models during both the blind and calibration phases. Among the MAAs, the MLR C method typically provided the closest match to measured daily ETa values, while the MLR A method was most accurate for maize yield across all sites and phases. The simple mean consistently outperformed the median at all sites. Therefore, MLR A and MLR C are recommended for averaging simulations of yield and ETa, respectively, when measured data is available. However, in the absence of observed ETa and yield data, the SMA method can be used to ensemble the yield and ETa simulations.

- **Individual Maize Model Performance:** No single maize model consistently performed best at all sites for simulating yield and daily ETa. Results indicate that fully calibrating the crop model, slightly improved the daily ETa simulation and moderately improved the yield estimates compared to the blind phase.
- **Phase Comparison for modeling averaging:** The performance of all MAAs improved slightly to moderately for daily ETa and yield from the blind phase to the calibrated phase across all sites.
- **Ensemble Member Models**: Using an ensemble of group maize models with different model structures slightly enhanced the accuracy of daily ETa and yield simulations at Group B in comparison to using an ensemble of all maize models.

These findings highlight the potential of MAAs to improve the precision of maize yield and daily ETa estimates, emphasizing the importance of using diverse model ensembles to achieve accurate agricultural predictions.

Credit authorship contribution statement

Viveka Nand: Conceptualization, Methodology, Software, Data analysis and interpretation, Visualisation, Writing–original draft**; Zhiming Qi:** Conceptualization, Methodology**,** Supervision; Writing – review and editing; (**Liwang Ma, Ward N. Smith**): Field study data processing**,** Writing – review and editing; (**Matthew J. Helmers, Chandra A. Madramootoo, Tiequan Zhang, Elizabeth Pattey, Virginia L. Jin, Thomas J. Trout, Andrew E. Suyker, Steven R. Evett, David K. Brauer, Gwen G. Coyle, Karen S. Copeland, Gary W. Marek, Paul D. Colaizzi**): Field study data processing; **Tobias KD Weber:** Supervision, Software, Writing – review and editing; **Bruce A. Kimball:** Data acquisition and aggregation of Maize AgMIP project; Software, Writing – review and editing; (**Daren Harmel, Kelly R. Thorp, Zoltán Barcza, Pasquale Garofalo, Antonio Trabucco, Michael van der Laan, Dennis Timlin**):Software, Writing – review and editing; (**Ziwei Li ,Jiaxin Wang, Qianjing Jiang, Haomiao Cheng, Kenneth J. Boote, Claudio Stockle, David K. Brauer, Gwen G. Coyle, Karen S. Copeland, Gary W. Marek, Paul D. Colaizzi, Marco Acutis, Seyyed Majid Alimagham, Sotirios Archontoulis, Faye Babacar, Zoltán Barcza, Bruno Basso, Patrick Bertuzzi, Julie Constantin, Massimiliano De Antoni Migliorati, Benjamin Dumont, Jean-Louis Durand, Nándor Fodor, Thomas Gaiser, Sebastian Gayler, Luisa Giglio, Robert Grant, Kaiyu Guan, Gerrit Hoogenboom, Soo-Hyung Kim, Isaya Kisekka, Jon Lizaso, Sara Masia, Huimin Meng, Valentina Mereu, Ahmed Mukhtar, Alessia Perego, Bin Peng,**

Eckart Priesack, Vakhtang Shelia, Richard Snyder, Afshin Soltani, Donatella Spano, Amit Srivastava, Aimee Thomson, Heidi Webber, Magali Willaume, Karina Williams, Domenico Ventrella, Michelle Viswanathan, Xu Xu, Wang Zhou): Software.

Declaration of competing interest

The authors state that they have no known financial conflicts of interest or personal relationships that could have influenced the work presented in this paper.

Data availability: Data will be made available by the corresponding author upon receiving a reasonable request.

Acknowledgments

We are grateful to the Ministry of Social Justice and Empowerment, Government of India (11015/48/2018-SCD-V), McGill University, and the Natural Sciences and Engineering Research Council of Canada (NSERC) for providing financial support for the first author to carry out this study.

References

Ahuja, L.R., Hanson, K.W., Rojas, K.W., Shaffer, M.J., 2000. Root Zone Water Quality Model. Modeling Management Effects on Water Quality and Crop Production. Water Resources Publications, Highlands Ranch, CO.

Aiolfi, M., Timmermann, A., 2006. Persistence in Forecasting Performance and Conditional Combination Strategies. *Journal of Econometrics* 135(1), 31–53.

Ajami, N.K., Duan, Q., Gao, X. and Sorooshian, S., 2006. Multimodel combination techniques for analysis of hydrological simulations: Application to distributed model intercomparison project results. *Journal of Hydrometeorology 7*(4), 755-768.

Akaike, H., 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control 19*(6), 716-723.

Albanito, F., McBey, D., Harrison, M., Smith, P., Ehrhardt, F., Bhatia, A., Bellocchi, G., Brilli, L., Carozzi, M., Christie, K. and Doltra, J., 2022. How modelers model: the overlooked social and human dimensions in model intercomparison studies. *Environmental Science & Technology*, *56*(18), pp.13485-13498.

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

Armstrong, J.S., 1989. Combining Forecasts: The End of the Beginning or the Beginning of the End? *International Journal of Forecasting* 5(4), 585–588.

Arsenault, R., Gatien, P., Renaud, B., Brissette, F., Martel, J., 2015. A comparative analysis of 9 multi-model averaging approaches in hydrological continuous streamflow simulation. *Journal of Hydrology* 529, 754–767. https://doi.org/10.1016/j.jhydrol.2015.09.001.

Asseng, S., Ewert, F., Rosenzweig, C., Jones, J.W., Hatfield, J.L., Ruane, A.C., Boote, K.J., Thorburn, P.J., Rötter, R.P., CaMAAsrano, D. and Brisson, N., 2013. Uncertainty in simulating wheat yields under climate change. *Nature Climate Change 3*(9), 827-832.

Bassu, S., Brisson, N., Durand, J., Boote, K., Lizaso, J., Jones, J.W., Rosenzweig, C., Ruane, A.C., Adam, M., Baron, C., Basso, B., Biernath, C., Boogaard, H., Conijn, S., Corbeels, M., Deryng, D., De Sanctis, G., Gayler, S., Grassini, P., Hatfield, J., Hoek, S., Izaurralde, C., Jongschaap, R., Kemanian, A.R., Kersebaum, K.C., Kim, S., Kumar, N.S., Makowski, D., Müller, C., Nendel, C., Priesack, E., Pravia, M.V., Sau, F., Shcherbak, I., Tao, F., Teixeira, E., Timlin, D., and Waha, K. *Global Change Biology* 20(7), 2301-2320.

Bates, J.M., Granger, C.W.J., 1969. The Combination of Forecasts. *Journal of the Operational Research Society* 20(4), 451–468.

Campbell, G.S. 1985. Soil Physics with BASIC, Elsevier, New York, New York. 150 pp.

Cheng, H., Shu, K., Qi, Z., Ma, L., Jin, V.L., Li, Y., Schmer, M.R., Wienhold, B.J. and Feng, S., 2021. Effects of residue removal and tillage on greenhouse gas emissions in continuous corn systems as simulated with RZWQM2. *Journal of Environmental Management* 285,112097.

Couëdel, A., Falconnier, G.N., Adam, M., Cardinael, R., Boote, K., Justes, E., Smith, W.N., Whitbread, A.M., Affholder, F., Balkovic, J. and Basso, B., 2024. Long-term soil organic carbon and crop yield feedbacks differ between 16 soil-crop models in sub-Saharan Africa. *European Journal of Agronomy*, *155*, p.127109.

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

Deb, P., Moradkhani, H., Han, X., Abbaszadeh, P., Xu, L. 2022. Assessing irrigation mitigating drought impacts on crop yields with an integrated modeling framework. *Journal of Hydrology*, 609, 127760.

Diks, C.G. and Vrugt, J.A., 2010. Comparison of point forecast accuracy of model averaging methods in hydrologic applications. *Stochastic Environmental Research and Risk Assessment* 24, 809-820.

Fang, Q., L. Ma, R.D. Harmel, Q. Yu, M.W. Sima, P.N.S. Bartling, R.W. Malone, B.T. Nolan, and J. Doherty. 2019. Uncertainty of CERES-Maize calibration under different irrigation strategies using PEST optimization algorithm. *Agronomy* 9(241), 1-17.

Farahani, H.J., DeCoursey, D.G., 2000. Potential evaporation and transpiration processes in the soil residue-canopy system. In: Ahuja, L.R., Rojas, K.W., Hanson, J.D., Shaffer, M.J., Ma, L. (Eds.), Root Zone Water Quality Model. Water Resources Publications, Highland Ranch, CO, pp. 51–80.

Fletcher, D.., 2018. *Why Model Averaging?* (pp. 1-29). Springer Berlin Heidelberg.

Gao, Y., Wallach, D., Hasegawa, T., Tang, L., Zhang, R., Asseng, S., Kahveci, T., Liu, L., He, J. and Hoogenboom, G., 2021. Evaluation of crop model prediction and uncertainty using Bayesian parameter estimation and Bayesian model averaging. *Agricultural and Forest Meteorology* 311, 108686.

Granger, C.W.J., Ramanathan, R., 1984. Improved methods of combining forecasts. *Journal of Forecasting* 3(2), 197-204.

He, D., Wang, E., Wang, J. and Robertson, M.J., 2017. Data requirement for effective calibration of process-based crop models. *Agricultural and Forest Meteorology* 234, 136-148.

Hoogenboom, G., Porter, C.H., Boote, K.J., Shelia, V., Wilkens, P.W., Singh, U., White, J.W., Asseng, S., Lizaso, J.I., Moreno, L.P. and Pavan, W., 2019. The DSSAT crop modeling ecosystem. In: *Advances in crop modelling for a sustainable agriculture* (pp. 173-216). Burleigh Dodds Science Publishing.

Holzworth, D.P., Huth, N.I., deVoil, P.G., Zurcher, E.J., Herrmann, N.I., McLean, G., Chenu, K., van Oosterom, E.J., Snow, V., Murphy, C. and Moore, A.D., 2014. APSIM–evolution towards a new generation of agricultural systems simulation. *Environmental Modelling & Software* 62, 327-350.

Huang, X., Huang, G., Yu, C., Ni, S. and Yu, L., 2017. A multiple crop model ensemble for improving broad-scale yield prediction using Bayesian model averaging. *Field Crops Research* 211, 114-124.

Ishaque, W., Osman, R., Hafiza, B.S., Malghani, S., Zhao, B., Xu, M., Ata-Ul-Karim, ST., 2023. Quantifying the impacts of climate change on wheat phenology, yield, and evapotranspiration under irrigated and rainfed conditions. *Agricultural Water Management* 275, 108017.

Jafarzadeh, A., Khashei-Siuki, A., and Pourreza-Bilondi, M., 2022. Performance assessment of model averaging techniques to reduce structural uncertainty of groundwater modeling. *Water Resources Management* 36(1), 353-377.

Jamieson, P. D., Porter, J. R., & Wilson, D. R. 1991. A test of the computer simulation model ARCWHEAT on wheat crops grown in New Zealand. *Field Crops Research* 27(4), 337-350.

Jiang, Q., Qi, Z., Madramootoo, C.A. and Singh, A.K., 2018. Simulating hydrologic cycle and crop production in a subsurface drained and sub-irrigated field in Southern Quebec using RZWQM2. *Computers and Electronics in Agriculture* 146, 31-42.

Jiang, Q., Qi, Z., Lu, C., Tan, C.S., Zhang, T., and Prasher, S.O., 2020. Evaluating RZ-SHAW model for simulating surface runoff and subsurface tile drainage under regular and controlled drainage with subirrigation in southern Ontario. *Agricultural water Management* 237, 106179.

Jones, J.W., Hoogenboom, G., Porter, C.H., Boote, K.J., Batchelor, W.D., L.A. Hunt, L.A., Wilkens, P.W., Singh, U., A.J. Gijsman, A.J., Ritchie, A.J., 2003. DSSAT Cropping System Model. *European Journal of Agronomy* 18, 235-265.

Jones, C.A. and Kiniry, J.R., 1986. *Ceres-maize; A simulation model of maize growth and development. eds* (No. 633.153 JON. CIMMYT.).

Keating, B.A., Carberry, P.S., Hammer, G.L., Probert, M.E., Robertson, M.J., Holzworth, D., Huth, N.I., Hargreaves, J.N.G., Meinke, H., Hochman, Z., McLean, G., Verburg, K., Snow, V., Dimes, J.P., Silburn, M., Wang, E., Brown, S., Bristow, K.L., Asseng, S., Chapman, S., McCown, R.L., Freebairn, D.M., Smith, C.J., 2003. An overview of APSIM: a model designed for farming systems simulation. European Journal of Agronomy 18, 267–288.

Kim, W., Iizumi, T., Nishimori, M., 2019. Global patterns of crop production losses associated with droughts from 1983 to 2009. *Journal of Applied Meteorology and Climatology* 58, 1233–1244.

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

Kimball, B.A., Thorp, K.R., Boote, K.J., Stockle, C., Suyker, A.E., Evett, S.R., Brauer, D.K., Coyle, G.G., Copeland, K.S., Marek, G.W., Colaizzi, P.D., Acutis, M., Alimagham, S., Archontoulis, S., Babacar, F., Barcza, Z., Basso, B., Bertuzzi, P., Constantin, J., De Antoni Migliorati, M., Dumont, B., Durand, J., Fodor, N., Gaiser, T., Garofalo, P., Gayler, S., Giglio, L., Grant, R., Guan, K., Hoogenboom, G., Jiang, Q., Kim, S., Kisekka, I., Lizaso, J., Masia, S., Meng, H., Mereu, V., Mukhtar, A., Perego, A., Peng, B., Priesack, E., Qi, Z., Shelia, V., Snyder, R., Soltani, A., Spano, D., Srivastava, A., Thomson, A., Timlin, D.J., Trabucco, A., Webber, H., Weber, T., Willaume, M., Williams, K., van der Laan, M., Ventrella, D., Viswanathan, M., Xu, X., and Zhou, W. 2023. Simulation of evapotranspiration and yield of maize: An inter-comparison among 41 maize models. *Agricultural and Forest Meteorology*. 333, 109396. doi.org/10.1016/j.agrformet.2023.109396.

Knipper, K., Anderson, M., Bambach, N., Melton, F., Ellis, Z., Yang, Y., Volk, J., McElrone, A.J., Kustas, W., Roby, M. and Carrara, W., 2024. A comparative analysis of OpenET for evaluating evapotranspiration in California almond orchards. *Agricultural and Forest Meteorology* 355, 110146.

Kothari, K., Battisti, R., Boote, K. J., Archontoulis, S.V., Confalone, A., Constantin, J., Cuadra, S.V., Debaeke, P., Faye, B., Grant, B., Hoogenboom, G., Jing, Q., van der Laan, M., da Silva, F. A.M., Marin, F.R., Nehbandani, A., Nendel, C., Purcell, L.C., Qian, B.D., Ruane, A.C., Schoving, C., Silva, E., Smith, W., Soltani, A., Srivastava, A., Vieira, N.A., Slone, S., and Salmeron, M. 2022. Are soybean models ready for climate change food impact assessments? *European Journal of Agronomy* 135, 15. https://doi.org/10.1016/j.eja.2022.126482.

Ma, L., Ahuja, L.R., Nolan, B.T., Malone, R.W., Trout, T.J. and Qi, Z., 2012. Root zone water quality model (RZWQM2): Model use, calibration, and validation. *Transactions of the ASABE 55*(4), 1425-1446.

Martre, P., Wallach, D., Asseng, S., Ewert, F., Jones, J.W., R¨ otter, R.P., Boote, K.J., Ruane, A.C., Thorburn, P.J., CaMAAsrano, D., Hatfield, J.L., Rosenzweig, C., Aggarwal, P.K., Angulo, C., Basso, B., Bertuzzi, P., Biernath, C., Brisson, N., Challinor, A.J., Doltra, J., Gayler, S., Goldberg, R., Grant, R.F., Heng, L., Hooker, J., Hunt, L.A., Ingwersen, J., Izaurralde, R.C., Kersebaum, K.C., Müller, C., Kumar, S. N., Nendel, C., O'Leary, G., Olesen, J.E., Osborne, T.M., Palosuo, T., Priesack, E., Ripoche, D., Semenov, M.A., Shcherbak, I., Steduto, P., Stockle, C.O., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Travasso, M., Waha, K., White, J.W., Wolf, J., 2015. Multimodel ensembles of wheat growth: many models are better than one. Global Change Biology 21, 911–925.

Neuman, S.P., 2003. Relationship between juxtaposed, overlapping, and fractal representations of multimodal spatial variability. *Water Resources Research* 39(8), 1-11.

Priestley, C.H.B. and Taylor, R.J., 1972. On the assessment of surface heat flux and evaporation using large-scale parameters. *Monthly Weather Review* 100(2), 81-92.

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

Qi, Z., Helmers, M.J., Malone, R.W. and Thorp, K.R., 2011. Simulating long-term impacts of winter rye cover crop on hydrologic cycling and nitrogen dynamics for a corn-soybean crop system. *Transactions of the ASABE 54*(5), 1575-1588.

Qi, Z., Ma, L., Bausch, W.C., Trout, T.J., Ahuja, L.R., Flerchinger, G.N. and Fang, Q., 2016. Simulating maize production, water and surface energy balance, canopy temperature, and water stress under full and deficit irrigation. *Transactions of the ASABE 59*(2), 623-633.

Ritchie, J.T., 1972. Model for predicting evaporation from a row crop with incomplete cover. *Water Resources Research* 8, 1204-1213.

Sándor, R., Ehrhardt, F., Grace, P., Recous, S., Smith, P., Snow, V., Soussana, J.F., Basso, B., Bhatia, A., Brilli, L. and Doltra, J., 2023. Residual correlation and ensemble modelling to improve crop and grassland models. *Environmental Modelling & Software*, *161*, p.105625.

Shamseldin, A.Y., O'Connor, K.M., Liang, G.C., 1997. Methods for combining the outputs of different rainfall–runoff models. *Journal of Hydrology* 197(1), 203–229.

Schwarz, G., 1978. Estimating the dimension of a model. *The annals of statistics*, 461-464.

Shuttleworth, W.J., and Wallace, J.S., 1985. Evaporation from sparse crops - an energy combination theory. *Quartarly Journal of the Royal* Meteorological *Society* 111, 839-855.

Singh, A. K., 2013. Water and nitrogen use efficiency of corn (*Zea mays* L.) under water table management. Ph.D. thesis, McGill University, Montreal, Canada.

Task Committee on Revision of Manual 70, 2016, April. Evaporation, evapotranspiration, and irrigation water requirements. American Society of Civil Engineers.

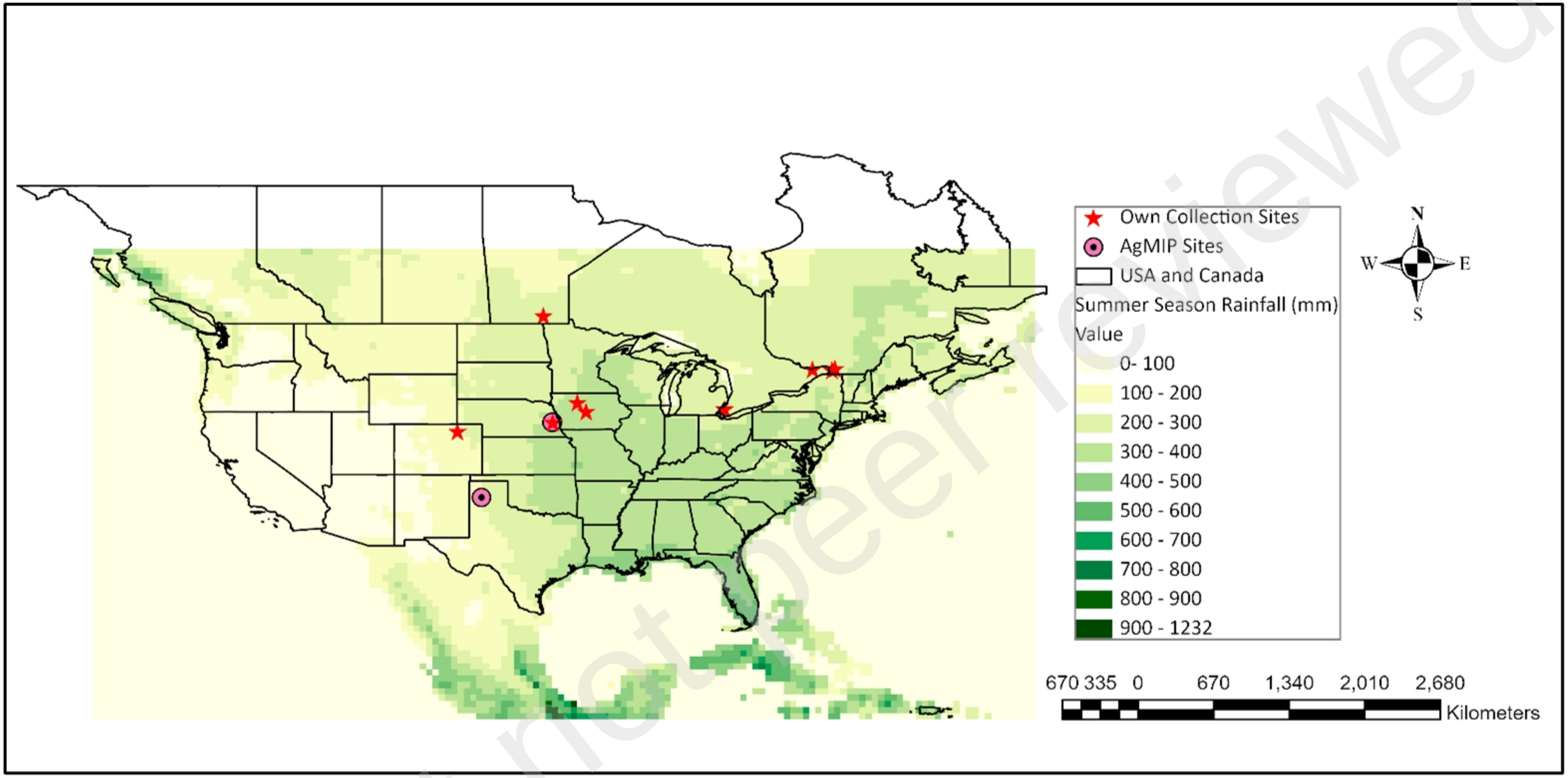
Uzoma, K.C., Smith, W., Grant, B., Desjardins, R.L., Gao, X., Hanis, K., Tenuta, M., Goglio, P. and Li, C., 2015. Assessing the effects of agricultural management on nitrous oxide emissions using flux measurements and the DNDC model. *Agriculture, Ecosystems & Environment 206*, 71-83.

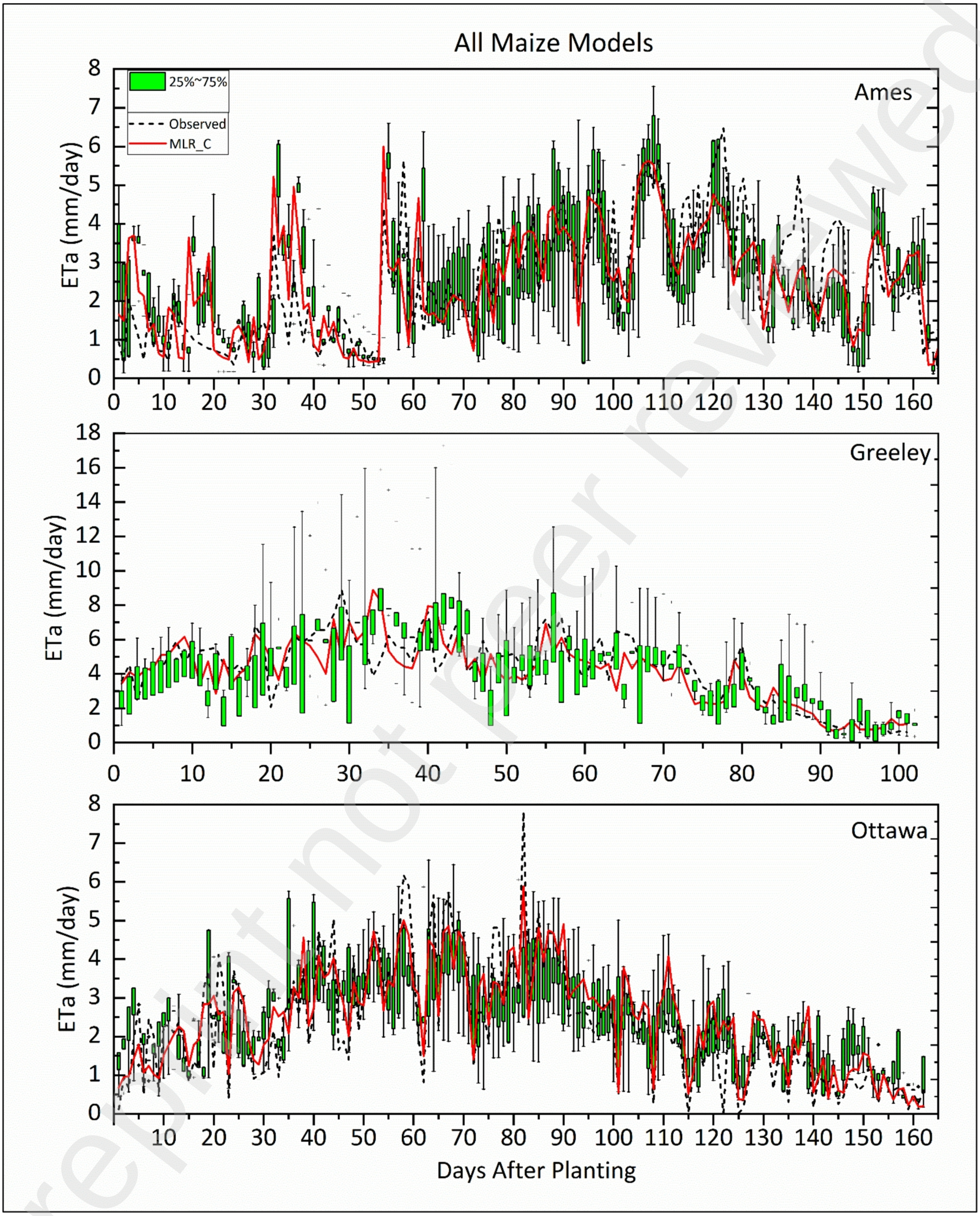
Wallach, D., Mearns, L.O., Ruane, A.C., Rötter, R.P. and Asseng, S., 2016. Lessons from climate modeling on the design and use of ensembles for crop modeling. *Climatic Change*, *139*, 551-564.

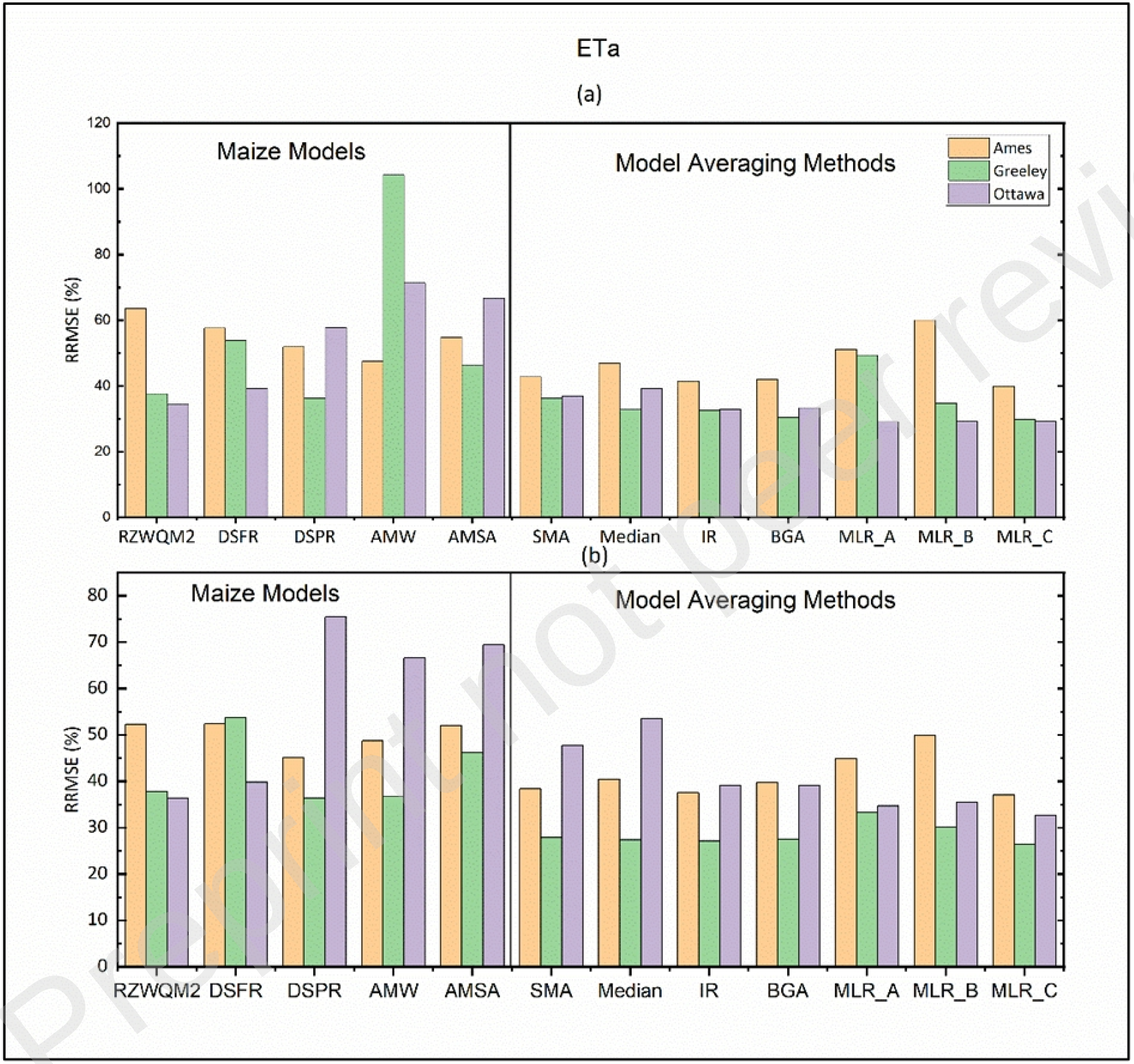
Wan, Y., Chen, J., Xu, C.-Y., Xie, P., Qi, W., Li, D. & Zhang, S. 2021 Performance dependence of multi-model combination methods on hydrological model calibration strategy and ensemble size. *Journal of Hydrology* 603, 127065.

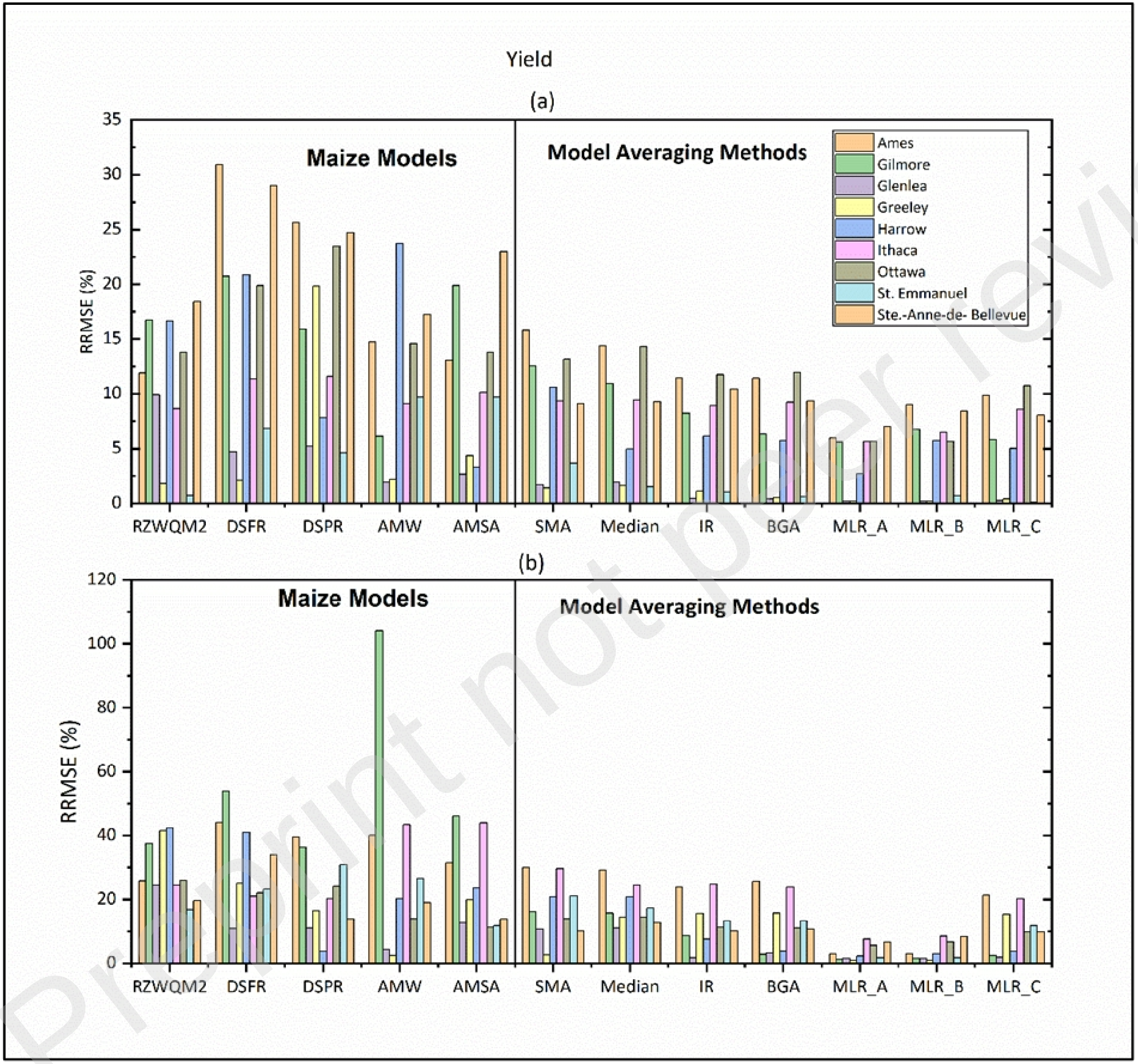
Yasin, M., Ahmad, A., Khaliq, T., Habib-ur-Rahman, M., Niaz, S., Gaiser, T., Ghafoor, I., Hassan, H.S.U., Qasim, M. and Hoogenboom, G., 2022. Climate change impact uncertainty assessment and adaptations for sustainable maize production using multi-crop and climate models. *Environmental Science and Pollution Research* 29, 1-22.

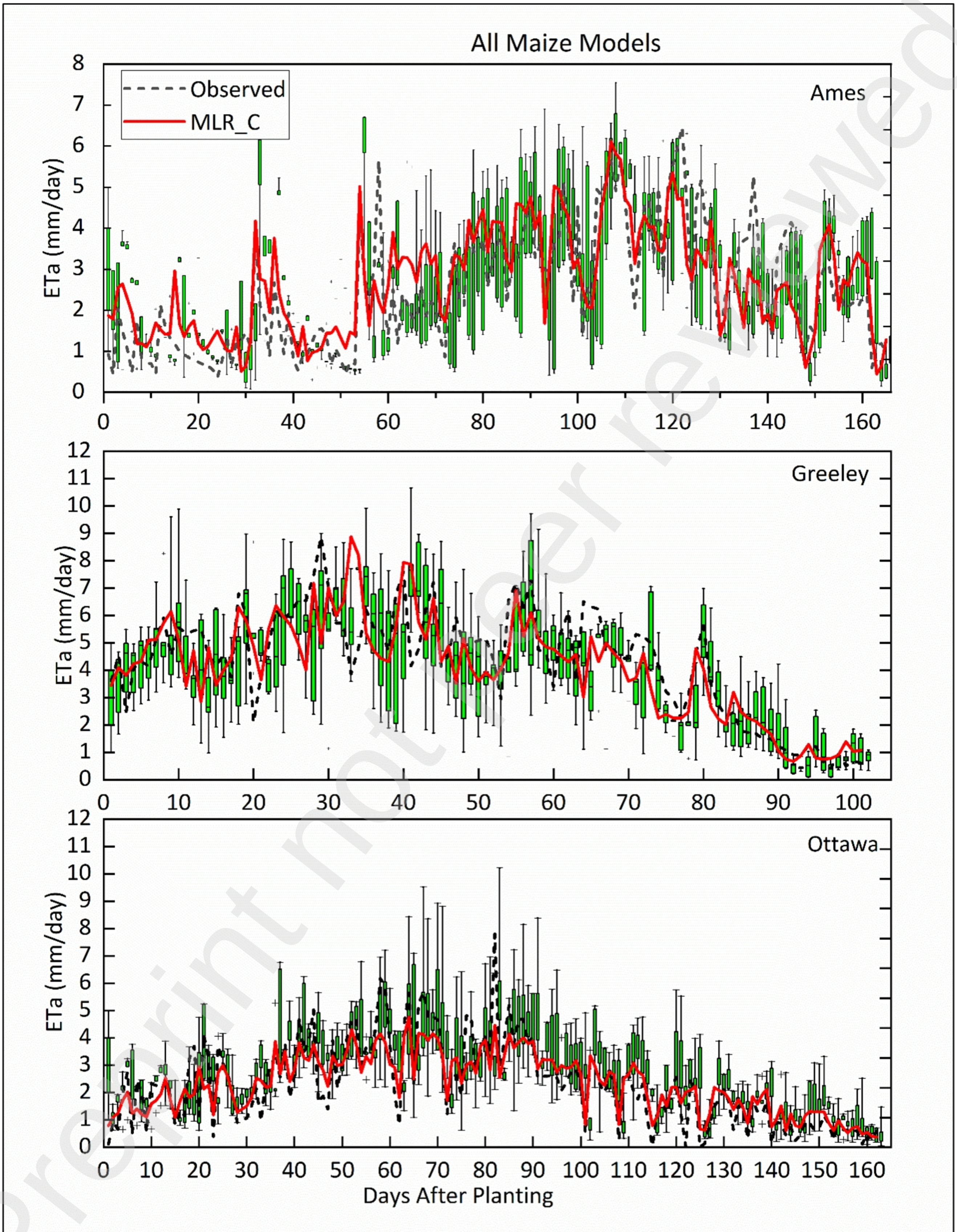
Zaherpour, J., Mount, N., Gosling, S.N., Dankers, R., Eisner, S., Gerten, D., Liu, X., Masaki, Y., Schmied, H.M., Tang, Q. and Wada, Y., 2019. Exploring the value of machine learning for weighted multi-model combination of an ensemble of global hydrological models. *Environmental Modelling & Software* 114, 12-128.


Own Collection Sites
AgMIP Sites
USA and Canada
Summer Season Rainfall (mm)
Value
0- 100
100 - 200
200 - 300
300 - 400
400 - 500
500 - 600
600 - 700
700 - 800
800 - 900
900 - 1232
N
W
E
S
670 335 0 670 1,340 2,010 2,680
Kilometers

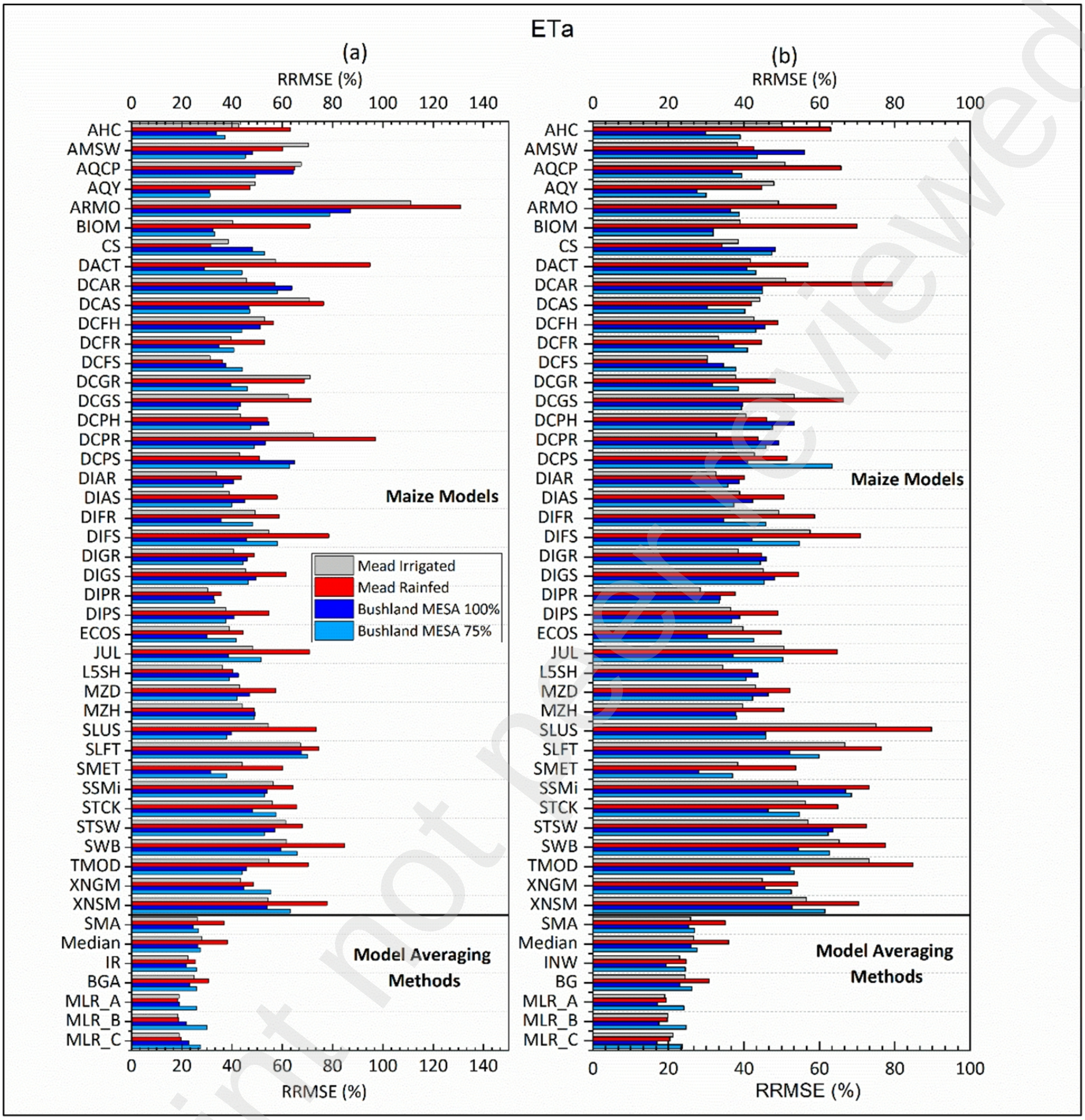

All Maize Models
25%~75%
Observed
MLR_C
Ames
Greeley
Ottawa
ETa (mm/day)
Days After Planting


ETa
(a)
Maize Models
Model Averaging Methods
Ames
Greeley
Ottawa
RRMSE (%)
RZWQM2
DSFR
DSPR
AMW
AMSA
SMA
Median
IR
BGA
MLR_A
MLR_B
MLR_C
(b)
Maize Models
Model Averaging Methods
RRMSE (%)

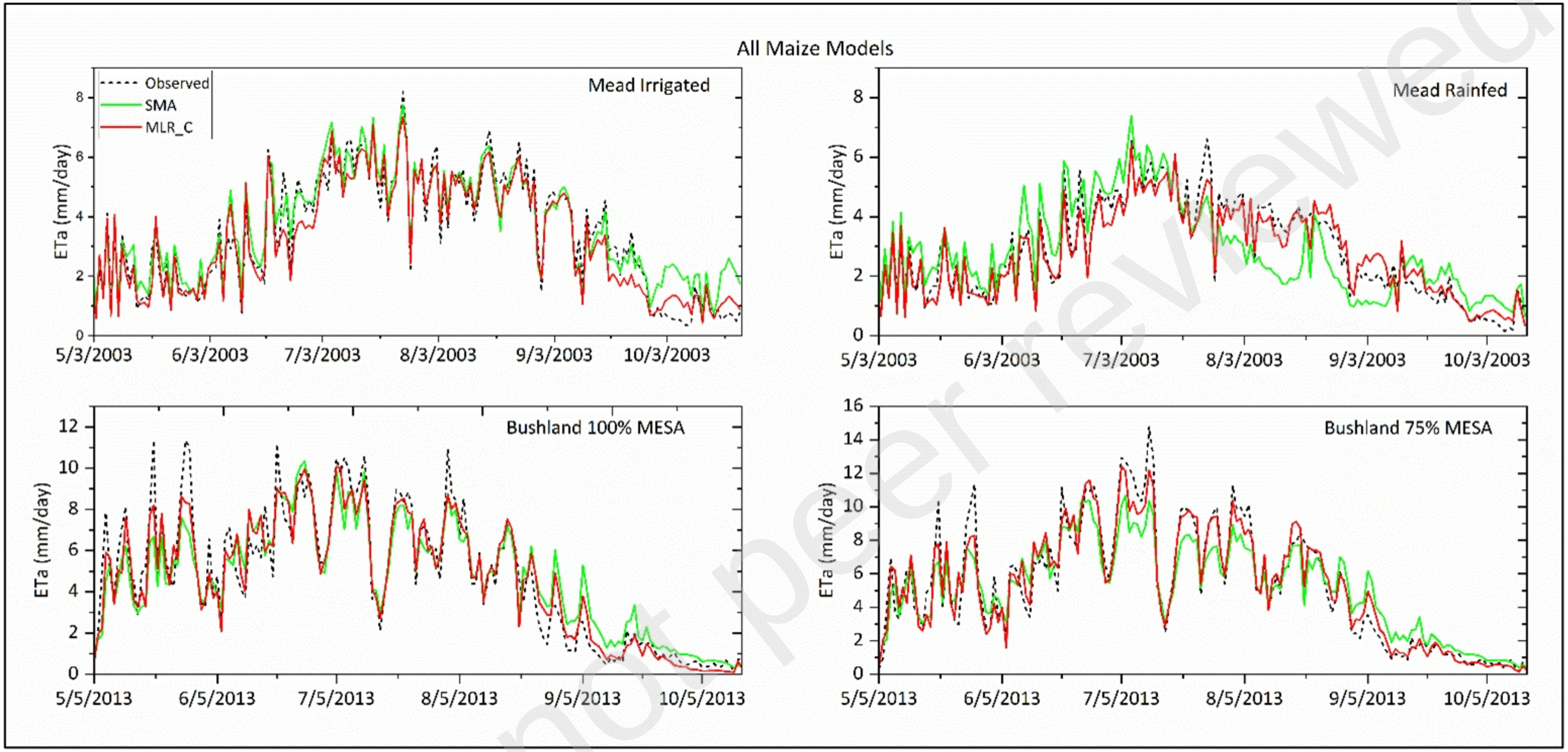

Yield
(a)
Maize Models
Model Averaging Methods
Ames
Gilmore
Glenlea
Greeley
Harrow
Ithaca
Ottawa
St. Emmanuel
Ste.-Anne-de- Bellevue
RRMSE (%)
0
5
10
15
20
25
30
35
RZWQM2
DSFR
DSPR
AMW
AMSA
SMA
Median
IR
BGA
MLR_A
MLR_B
MLR_C
(b)
Maize Models
Model Averaging Methods
RRMSE (%)
0
20
40
60
80
100
120
RZWQM2
DSFR
DSPR
AMW
AMSA
SMA
Median
IR
BGA
MLR_A
MLR_B
MLR_C

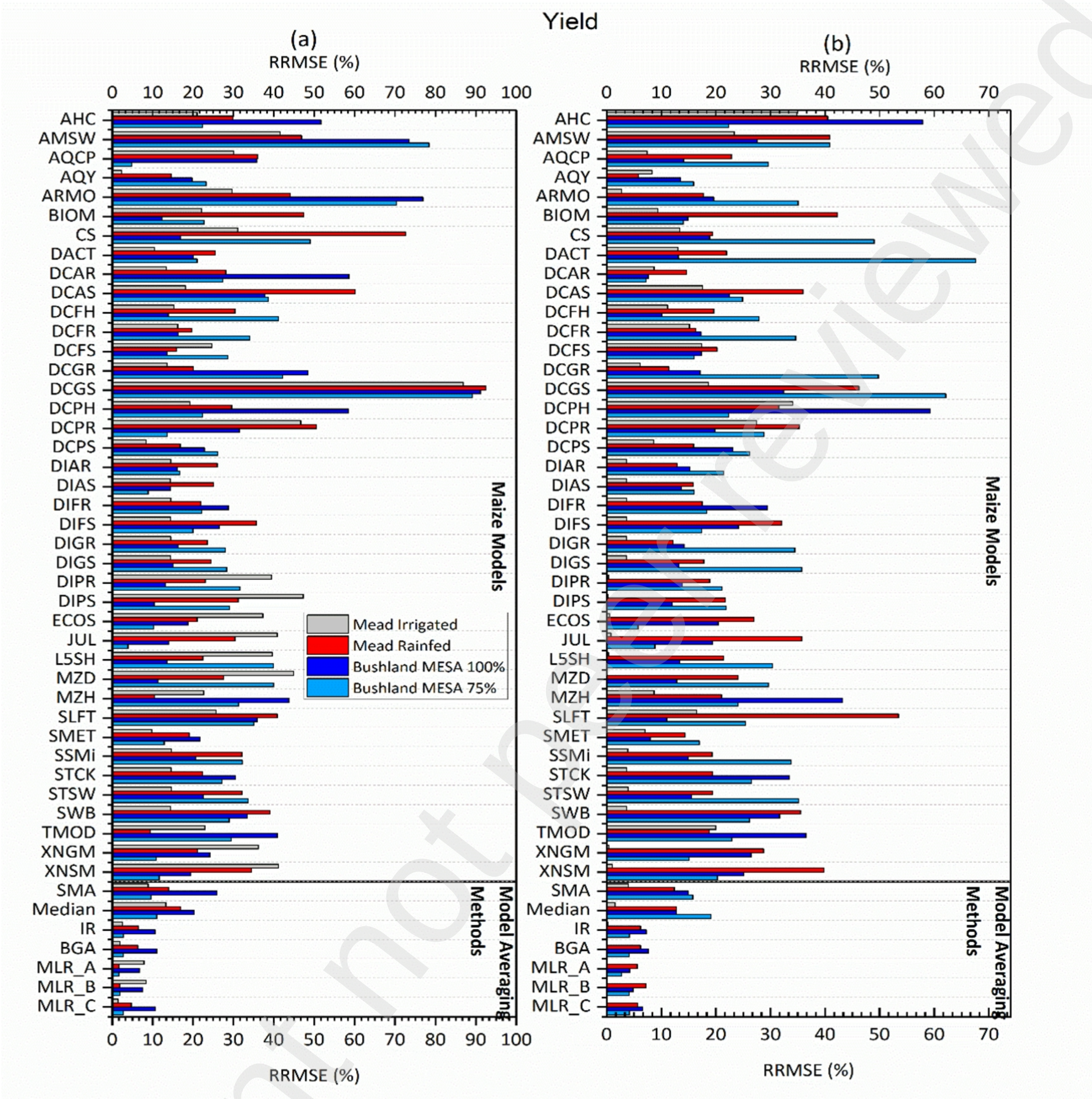


All Maize Models
Observed
MLR_C
Ames
Greeley
Ottawa
ETa (mm/day)
Days After Planting



ETa
(a)
RRMSE (%)
0 20 40 60 80 100 120 140
AHC
AMSW
AQCP
AQY
ARMO
BIOM
CS
DACT
DCAR
DCAS
DCFH
DCFR
DCFS
DCGR
DCGS
DCPH
DCPR
DCPS
DIAR
DIAS
DIFR
DIFS
DIGR
DIGS
DIPR
DIPS
ECOS
JUL
L5SH
MZD
MZH
SLUS
SLFT
SMET
SSMi
STCK
STSW
SWB
TMOD
XNGM
XNSM
SMA
Median
IR
BGA
MLR_A
MLR_B
MLR_C
Maize Models
Model Averaging Methods
Mead Irrigated
Mead Rainfed
Bushland MESA 100%
Bushland MESA 75%
(b)
RRMSE (%)
0 20 40 60 80 100
INW
BG


All Maize Models
Observed
SMA
MLR_C
Mead Irrigated
Mead Rainfed
Bushland 100% MESA
Bushland 75% MESA
ETa (mm/day)
5/3/2003
6/3/2003
7/3/2003
8/3/2003
9/3/2003
10/3/2003
5/5/2013
6/5/2013
7/5/2013
8/5/2013
9/5/2013
10/5/2013



Yield
(a)
(b)
RRMSE (%)
0 10 20 30 40 50 60 70 80 90 100
0 10 20 30 40 50 60 70
AHC
AMSW
AQCP
AQY
ARMO
BIOM
CS
DACT
DCAR
DCAS
DCFH
DCFR
DCFS
DCGR
DCGS
DCPH
DCPR
DCPS
DIAR
DIAS
DIFR
DIFS
DIGR
DIGS
DIPR
DIPS
ECOS
JUL
L5SH
MZD
MZH
SLFT
SMET
SSMi
STCK
STSW
SWB
TMOD
XNGM
XNSM
SMA
Median
IR
BGA
MLR_A
MLR_B
MLR_C
Maize Models
Model Averaging Methods
Mead Irrigated
Mead Rainfed
Bushland MESA 100%
Bushland MESA 75%

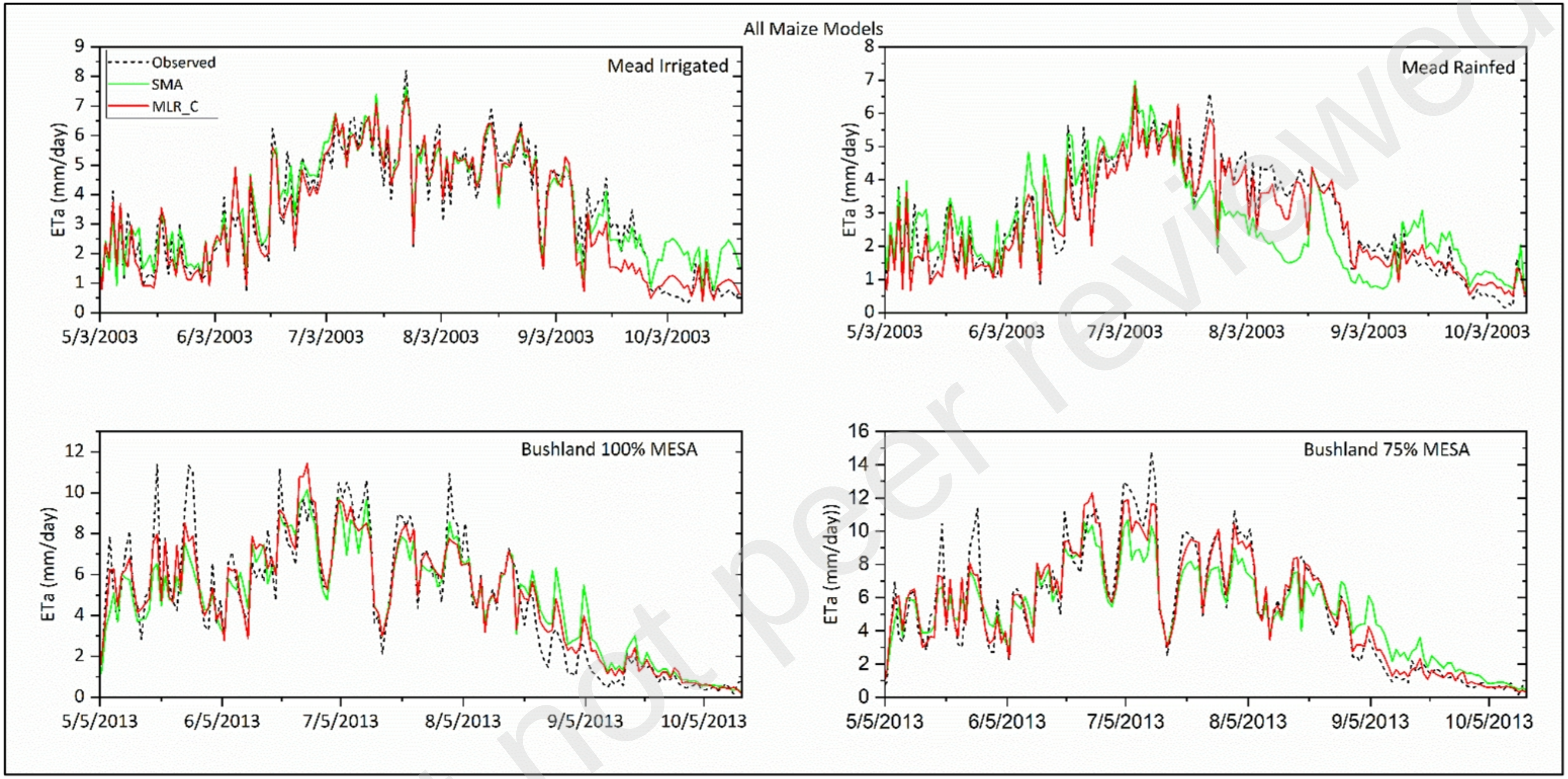

All Maize Models
Observed
SMA
MLR_C
Mead Irrigated
ETa (mm/day)
5/3/2003
6/3/2003
7/3/2003
8/3/2003
9/3/2003
10/3/2003
Mead Rainfed
Bushland 100% MESA
5/5/2013
6/5/2013
7/5/2013
8/5/2013
9/5/2013
10/5/2013
Bushland 75% MESA
ETa (mm/day))

Fig.1 Locations of crop field sites in the USA and Canada (own collection sites referred to as Group A sites, and AgMIP sites marked as Group B sites).
(Source: http://drought.memphis.edu/naspa/CompReconRange.aspx)

Fig.2. Box plots of daily simulated evapotranspiration (ETa) of the corn season 2006, 2010, and 2006 at Group A sites (Ames, Greeley, and Ottawa) respectively. Observed daily ETa values and the MLR C model averaging method derived daily ETa values from the 5 maize models are also presented. The simulated outputs of the blind phase where the model was set up using in-situ measured data and no calibration was done.

Fig. 3. RRMSE between the measured and simulated daily ETa across crop models and model averaging methods under blind (a) and calibration (b) phases at Group A sites.

Fig. 4. RRMSE between the measured and simulated maize yield across maize models and model averaging methods under blind (a) and calibration (b) phases Group A sites.

Fig.5. Box plots of daily simulated evapotranspiration (ETa) of the corn season 2006, 2010, and 2006 at Group A sites (Ames, Greeley, and Ottawa) respectively. Observed daily ETa values and the MLR C model averaging method derived daily ETa values from the 5 maize models are also presented. The simulated outputs of the calibrated phase where fully calibrated using crop phenology dates, LAI, soil moisture, ETa and yield data.

Fig.6. RRMSE between the measured and simulated daily crop evapotranspiration (ETa) across maize models and model averaging methods at Group B sites under blind (a) and calibration phase (b).

Fig.7. A comparison of measured daily ETa simulations and an ensemble of daily ETa simulations of all maize models using SMA and MLR C averaging methods at Group B sites under blind phase.

Fig.8. RRMSE between the measured and simulated maize yield across maize models and model averaging methods at Group B sites under blind (a) and calibration phase (b).

Fig.9. A comparison of measured daily ETa simulations and an ensemble of daily ETa simulations of all maize models using SMA and MLR C model averaging methods at Group B sites under the calibration phase.

Table 1. Details of selected crop field sites and corresponding soil type, average rainfall,and average temperature during the growing season (April-October).

Name	Country	ProvinceState	Lat	Long	Soil Type	Growing season climatic parameters: Rainfall (mm)	Growing season climatic parameters: Mean Temp (°C)	Modeled Component	Sources
Group A sites									
Ames	USA	Iowa	42.02	-93.75	Loam	536.37	18.62	Yield and AET	Kimbal et al., 2019
Gilmore	USA	Iowa	42.73	-94.45	Clay Loam	559.35	17.47	Yield	Qi et al.,2011
Glenlea	Canada	Manitoba	49.64	-97.16	Clay	399.00	14.10	Yield	Uzoma et. al.,2015
Greeley	USA	Colorado	40.44	-104.00	Loamy Sand	191.00	16.50	Yield and AET	Qi et al.,2016
Harrow	Canada	Ontario	42.22	-82.73	Clay Loam	505.93	18.21	Yield	Jiang et al.,2020
Ithaca	USA	Nebraska	41.16	-96.41	Silty Loam	592.36	10.40	Yield	Cheng et al., 2021
Ottawa	Canada	Ontario	45.38	-75.72	Loam	530.80	16.19	Yield and AET	Crépeau et al.,2021
St. Emmanuel	Canada	Québec	45.32	-74.17	Clay Loam	578.87	16.35	Yield	Singh, 2013
Ste.-Anne-de-Bellevue	Canada	Québec	45.43	-73.93	Loamy Sand	580.52	16.27	Yield	Jiang et al.,2022
Group B Sites									
Bushland	USA	Texas	35.18	-102.09	Silty Clay	350	22.80	Yield and AET	Kimbal et al., 2023
Mead Rainfed	USA	Nebraska	41.17	-96.43	Silty Loam	592	19.90	Yield and AET	Kimbal et al., 2023
Mead Irrigated	USA	Nebraska	41.16	-96.47	Silty Loam	592	19.90	Yield and AET	Kimbal et al., 2023

Table 2. A comparison of RRMSE between the measured daily ETa and ensembled daily ETa of all maize models and group maize models using seven model averaging methods at Group A sites under the Blind and Calibrated Phase.

Averaging Approaches	Blind						Calibrated					
	All Models			Group Models			All Models			Group Models		
	Ames	Greeley	Ottawa	Ames	Greeley	Ottawa	Ames	Greeley	Ottawa	Ames	Greeley	Ottawa
SMA	42.8	36.2	44.9	45.9	45.8	37	38.4	27.8	38.3	39.4	29.2	35.7
Median	47.0	32.8	49.7	45.9	45.8	39.3	40.5	27.4	38.1	46.2	29.9	35.7
IR	41.6	32.6	37.9	44.0	37.2	33	37.5	27.2	35.2	38.8	27.4	30.4
BGA	42.0	30.4	37.7	44.6	35.5	33.4	39.8	27.5	34.6	38.7	27.6	30.9
MLR A	51.2	49.5	34.7	47.7	39.2	29.1	44.9	33.4	29.8	47.4	33.8	28.7
MLR B	60.1	34.8	35.5	49.2	34.9	29.4	49.9	30.2	29.7	48.8	36.0	28.4
MLR C	40.0	29.8	32.4	43.9	35.0	29.2	37.1	26.4	32.3	38.0	27.3	29.2

Table 3. A comparison of RRMSE between the measured maize yield and ensembled maize yield of all maize models and group maize models using seven model averaging methods at Group A sites under the Blind and Calibrated Phase.

Averaging Approaches	Blind																	
	All Models									Group Models								
	Ames	Gilmore	Glenlea	Greeley	Harrow	Ithaca	Ottawa	St Emmanuel	Ste Anne De Bellevue	Ames	Gilmore	Glenlea	Greeley	Harrow	Ithaca	Ottawa	St Emmanuel	Ste Anne De Bellevue
SMA	29.9	16.2	10.8	2.6	20.9	29.6	14.0	21.1	10.1	29.8	16.2	10.1	6.8	28.0	28.5	16.2	17.1	12.2
Median	29.2	15.8	11.1	14.4	20.8	24.5	14.4	17.4	12.9	31.2	13.3	9.5	15.5	37.2	24.5	16.4	15.7	13.1
IR	24.0	8.7	1.7	15.7	7.7	24.9	11.3	13.3	10.2	25.3	10.8	1.9	7.6	17.8	25.2	15.1	13.3	10.2
BGA	25.7	2.8	3.3	15.8	3.9	24.0	11.1	13.3	10.7	25.1	2.7	2.6	3.0	17.0	24.2	15.0	13.3	10.9
MLR A	2.9	1.2	1.6	1.0	2.2	7.7	5.7	1.7	6.6	3.4	1.2	1.8	7.8	1.2	8.1	11.8	1.6	7.4
MLR B	3.1	1.6	1.5	1.0	3.0	8.6	6.8	1.7	8.4	3.6	2.4	1.8	7.8	1.9	9.3	13.5	1.6	8.2
MLR C	21.3	2.5	1.9	15.5	3.9	20.2	9.9	11.8	9.9	23.4	2.5	1.8	6.7	17.0	21.0	14.0	11.8	10.0
	Calibrated																	
	All Models									Group Models								
SMA	15.8	12.5	1.7	1.4	10.6	9.4	13.2	3.7	9.1	15.0	10.0	3.5	1.3	15.28	9.54	12.2	2.60	10.6
Median	14.4	10.9	2.0	1.6	5.0	9.4	14.3	1.5	9.3	13.9	8.7	4.0	3.2	20.86	9.28	14.6	0.72	12.2
IR	11.5	8.2	0.5	1.1	6.1	9.0	11.7	1.1	10.5	11.8	7.2	1.7	4.0	11.76	9.21	11.4	3.20	10.8
BGA	11.4	6.3	0.4	0.5	5.7	9.2	11.9	0.6	9.4	11.2	5.4	1.1	4.0	11.74	9.39	11.7	0.71	11.4
MLR A	5.9	5.6	0.2	0.2	2.7	5.7	5.7	0.1	7.0	6.3	2.4	0.2	0.1	1.70	7.84	8.5	0.08	8.6

MLR B	9.0	6.7	0.2	0.2	5.7	6.5	5.7	0.7	8.5	6.9	4.5	0.2	0.1	2.88	8.88	11.3	0.23	9.8
MLR C	9.9	5.8	0.2	0.4	5.0	8.6	10.7	0.1	8.1	10.9	5.4	0.6	4.0	11.94	9.00	10.7	0.93	9.7

Table 4. Comparison of RRMSE between (a) measured daily ETa and ensembled daily ETa, and (b) measured maize yield and ensembled maize yield for all maize models and group maize models using seven model averaging methods at Group B sites under Blind and Calibrated Phases.

	Daily ETa (a)															
Averaging Approaches	Blind								Calibrated							
	All Models				Group Models				All Models				Group Models			
	Mead Irrigated	Mead Rainfed	Bushland 100% MESA	Bushland 75% MESA	Mead Irrigated	Mead Rainfed	Bushland 100% MESA	Bushland 75% MESA	Mead Irrigated	Mead Rainfed	Bushland 100% MESA	Bushland 75% MESA	Mead Irrigated	Mead Rainfed	Bushland 100% MESA	Bushland 75% MESA
SMA	26.1	36.8	24.6	26.8	24.4	34.3	23.5	25.7	25.9	35.0	25.4	26.9	23.8	33.5	24.8	26.1
Median	28.0	38.1	26.4	27.5	24.4	34.6	25.7	27.2	26.7	35.9	26.0	27.6	23.6	33.8	25.6	27.1
IR	22.5	25.4	21.7	25.8	21.2	24.0	20.3	24.9	22.9	24.7	19.4	24.5	20.8	23.9	18.9	24.4
BGA	25.0	30.6	23.3	25.9	22.7	27.4	22.1	25.4	24.4	30.7	23.0	26.2	22.4	28.6	22.7	25.9
MLR A	18.4	18.7	21.8	30.0	18.6	18.1	19.5	30.3	19.0	19.4	17.1	24.1	19.4	19.4	17.0	24.0
MLR B	18.9	19.7	22.9	27.4	18.8	18.3	21.6	27.9	19.8	19.7	17.4	24.7	19.9	20.0	17.4	25.0
MLR C	18.9	18.5	19.0	25.9	18.9	18.5	16.9	22.2	21.2	20.4	17.0	23.7	19.7	20.5	16.2	21.5
	Seasonal Yield (b)															
SMA	8.9	14.0	26.0	9.6	7.4	11.7	28.3	11.8	4.0	12.4	15.0	15.8	5.1	9.8	14.9	16.6
Median	13.3	17.0	20.3	11.0	10.0	16.0	26.3	11.9	1.6	12.8	12.8	19.1	2.0	9.3	12.7	16.8
IR	2.6	6.5	10.7	2.8	3.8	6.5	10.1	16.8	0.2	6.2	7.3	4.2	0.4	7.3	7.2	4.2
BGA	2.0	6.4	11.0	2.8	2.7	6.4	10.2	15.8	0.1	6.2	7.7	4.1	0.3	6.8	7.5	4.1
MLR A	7.9	1.6	6.8	1.7	2.0	2.4	8.0	8.3	0.0	5.6	4.2	2.8	0.1	3.6	3.5	3.4
MLR B	8.4	1.9	7.6	1.9	2.4	4.2	8.3	10.1	0.1	7.2	4.8	4.1	0.3	5.9	5.9	4.1
MLR C	1.5	4.7	10.7	2.8	2.7	4.7	9.1	15.5	0.1	5.7	6.6	4.2	0.2	5.9	6.6	4.2

Table 5. Average RRMSE between measured daily ETa and maize yield, and ensembled daily ETa and maize yield for all models and group models at Group A sites for both the blind and calibration phases.

	Daily ETa					Seasonal Yield				
	Blind		Calibrated		Overall	Blind		Calibrated		Overall
	All Models	Group Models	All Models	Group Models		All Models	Group Models	All Models	Group Models	
SMA	38.7	42.4	37.1	35.6	38.4	17.3	18.3	8.6	8.9	13.3
Median	39.7	42.4	39.2	38.1	39.9	17.8	19.6	7.6	9.7	13.7
IR	35.7	37.2	34.2	33.8	35.2	13.1	14.1	6.6	7.9	10.4
BGA	35.3	37.0	35.0	33.6	35.2	12.3	12.6	6.2	7.4	9.6
MLR A	36.5	39.0	35.5	34.7	36.4	3.4	4.9	3.7	4.0	4.0
MLR B	36.0	38.3	35.2	34.5	36.0	4.0	5.6	4.8	5.0	4.8
MLR C	33.0	36.0	32.0	32.5	33.4	10.8	12.0	5.4	7.0	8.8
Mean	36.4	38.9	35.4	34.7	36.4	11.2	12.5	6.1	7.1	9.2

Table 6. Average RRMSE between measured daily ETa and yield, and ensembled daily ETa and yield for all models and group models at Group B sites for both the blind and calibration phases.

Averaging approaches	Daily ETa					Yield				
	Blind		Calibrated		Overall	Blind		Calibrated		Overall
	All Models	Group Models	All Models	Group Models		All Models	Group Models	All Models	Group Models	
SMA	28.6	27.0	28.3	27.0	27.7	14.6	14.8	11.8	11.6	13.2
Median	30.0	27.9	29.0	27.5	28.6	15.4	16.1	11.6	10.2	13.3
IR	23.9	22.6	22.9	22.0	22.8	5.6	9.3	4.5	4.8	6.0
BGA	26.2	24.4	26.1	24.9	25.4	5.5	8.8	4.5	4.7	5.9
MLR A	22.2	22.2	22.1	21.5	22.0	4.5	5.2	3.2	3.5	4.1
MLR B	22.2	22.2	21.9	21.5	22.0	4.9	6.3	4.0	4.2	4.9
MLR C	20.6	19.1	20.6	19.5	19.9	4.9	8.0	4.1	4.2	5.3
Mean	24.8	23.6	24.4	23.4	24.1	7.9	9.8	6.2	6.2	7.5