

Disentangling sources of errors in simulated wheat evapotranspiration

Chenzhi Wang¹, Diane Cooke¹, Xinxin Chen¹, Jeffrey White², Katja Trachte³, Pierre Martre ⁴,
 Senthold Asseng ⁵, Frank Ewert^{1,6}, Bruce Kimball⁷, Gerrit Hoogenboom ², Steven Evett⁸,
 André Chanzy⁹, Sébastien Garrigues¹⁰, Albert Olioso^{9,11}, Karen Copeland⁸, Jean Steiner¹²,
 Davide Cammarano¹³, Yi Chen^{14,15}, Marianne Crépeau¹⁶, Roberto Ferrise¹⁷, Loic Manceau ⁴,
 Thomas Gaiser ⁶, Yujing Gao^{2,18}, Sebastian Gayler¹⁹, Jose Guarin^{20,21}, Tony Hunt²², Guillaume Jégo¹⁶,
 Gloria Padovan¹⁷, Elizabeth Pattey²³, Dominique Ripoche²⁴, Alfredo Rodríguez²⁵,
 Margarita Ruiz Ramos²⁶, Vakhtang Shelia², Amit Srivastava^{1,6}, Iwan Supit²⁷, Fulu Tao^{13,14}, Kelly Thorp²⁸,
 Michelle Viswanathan²⁹, Tobias Karl David Weber^{30,31}, Heidi Webber ^{1,3*}

¹Leibniz Centre for Agricultural Landscape Research (ZALF), Eberswalder Straße 84, Müncheberg 15374, Germany

²Agricultural and Biological Engineering Department, University of Florida, 1741 Museum Road, Bldg. 474, Gainesville, FL 32611-0570, United States

³Institute of Environmental Science and Environmental Technology, Brandenburg University of Technology Cottbus-Senftenberg, Burger Ch 2, Cottbus 03044, Germany

⁴LEPSE, Univ Montpellier, INRAE, Institut Agro Montpellier, 2 place Pierre Viala, Montpellier Cedex 02 34060, France

⁵School of Life Sciences, Department of Life Science Engineering, Chair of Digital Agriculture, HEF World Agricultural Systems Center, Technical University of Munich, Liesel-Beckmann-Str. 2, Freising 85354, Germany

⁶Crop Science, Institute of Crop Science and Resource Conservation, University of Bonn, Friedrich-Ebert-Allee 144, Bonn 53115, Germany

⁷Independent Researcher (Retired), United States

⁸USDA Agricultural Research Service, Conservation and Production Research Laboratory, 2300 Experiment Station Road, Bushland, TX 79012, United States

⁹UMR EMMAH, INRAE-Avignon Université, 228 route de l'Aérodrome, Domaine Saint-Paul, Site Agroparc, Avignon Cedex 9 84914, France

¹⁰European Centre for Medium-Range Weather Forecasts (ECMWF), Shinfield Park, Reading RG2 9AX, United Kingdom

¹¹URFM, INRAE, 228 route de l'Aérodrome, Domaine Saint-Paul, Site Agroparc, Avignon Cedex 9 84914, France

¹²Department of Agronomy, Kansas State University, 1712 Claflin Road, Manhattan, KS 66506, United States

¹³Department of Agroecology, iClimate, CBIO, Aarhus University, Blichers Allé 20, Tjele 8830, Denmark

¹⁴College of Resources and Environment, University of Chinese Academy of Sciences, No. 1 Yanqihu East Rd, Huairou District, Beijing 100048, China

¹⁵Key Laboratory of Land Surface Pattern and Simulation, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, 11A Datun Road, Chaoyang District, Beijing 100101, China

¹⁶Science and Technology Branch, Agriculture and Agri-Food Canada, 2560 Hochelaga Boulevard, Québec City, QC G1V 2J3, Canada

¹⁷Department of Agriculture, Food, Environment and Forestry (DAGRI), University of Florence, Piazzale delle Cascine 18, Florence 50144, Italy

¹⁸Appalachian Laboratory, University of Maryland Center for Environmental Science, 301 Braddock Road, Frostburg, MD 21532, United States

¹⁹WESS - Water and Earth System Science Competence Cluster, University of Hohenheim, Schloß Hohenheim 1, Stuttgart 70599, Germany

²⁰NASA Goddard Institute for Space Studies, 2880 Broadway, New York, NY 10025, United States

²¹Autonomic Integra LLC, 6550 Rock Spring Drive, Suite 600, Bethesda, MD 20817, United States

²²University of Guelph, 50 Stone Road East, Guelph, ON N1G 2W1, Canada

²³Science and Technology Branch, Agriculture and Agri-Food Canada, 960 Carling Avenue, Ottawa, Canada (Retired)

²⁴AgroClim, INRAE Domaine Saint-Paul, 228 route de l'Aérodrome, Site Agroparc, Avignon Cedex 9 84914, France

²⁵Department of Economic Analysis and Finances, University of Castilla-La Mancha, Toledo 45071, Spain

²⁶CEIGRAM, Technical University of Madrid, C/ Senda del Rey 13, Madrid 28040, Spain

²⁷Earth Systems and Global Change Group, Wageningen University & Research, Droevendaalsesteeg 3a, Wageningen 6708 PB, The Netherlands

²⁸USDA Agricultural Research Service, Grassland Soil and Water Research Laboratory, 808 East Blackland Road, Temple, TX 76502, United States

²⁹Julius Kühn Institute (JKI), Institute for Strategies and Technology Assessment, Stahnsdorfer Damm 81, Kleinmachnow 14532, Germany

³⁰Section of Soil Science and Agrohydrology, University of Kassel, Nordbahnhofstraße 1A, Witzenhausen 37213, Germany

³¹Biogeophysics, Institute of Soil Science and Land Evaluation, University of Hohenheim, Emil-Wolff-Straße 27, Stuttgart 70599, Germany

*Corresponding author. Agricultural Landscape Systems, Leibniz Centre for Agricultural Landscape Research (ZALF), Eberswalder Straße 84, Müncheberg 15374, Germany. E-mail: heidi.webber@zalf.de

Abstract

Accurate simulation of crop evapotranspiration (ET) is critical for water management and for understanding the impacts of climate variability and change. Yet evidence on model skill in simulating ET remains limited, and the sources of error remain poorly understood. Daily simulations from 17 wheat crop models were evaluated against field observations from two sites in semi-arid and Mediterranean environments under water-limited and irrigated conditions. Absolute ET errors were larger under irrigation, but this contrast was substantially attenuated after normalization by observed ET. The larger absolute errors under irrigation were mainly associated with a greater lack-of-correlation component rather than stronger systematic bias, indicating greater day-to-day mismatch between simulations and observations. Differences among ET-demand approaches also tended to be more evident under irrigation, with Priestley–Taylor-based models tending to produce larger errors, although the broad method classes also differed in other model components. Mixed-effects attribution showed that the conditional associations between ET error and climatic factors differed between the two water-supply datasets. Under irrigation, ET simulation error was associated with vapour pressure deficit (VPD), wind speed, radiation, and the component of maximum temperature not linearly explained by VPD, whereas under water-limited conditions meteorological attribution was weaker and observed soil–water status provided additional information. Root-zone soil moisture provided greater model support than shallow or deep soil moisture and modified the association between VPD and ET error. These findings identify atmospheric-demand representation and root-zone water availability as priorities for improving the evaluation and development of crop ET simulations.

Keywords: crop models, evapotranspiration, irrigated and water-limited conditions, soil moisture, error decomposition, vapour pressure deficit

1. Introduction

Evapotranspiration (ET) is a central flux in the soil–plant–atmosphere continuum, governing surface energy partitioning while also linking the terrestrial water, carbon, and nutrient cycles (Fisher et al. 2017, Piao et al. 2020, Lian et al. 2021) to climatic conditions (Seneviratne et al. 2010). At the global scale, ET returns more than 60% of land precipitation to the atmosphere (Oki and Kanae 2006) and consumes about half of net radiation at the land surface, underscoring its fundamental role in Earth system functioning (Katul et al. 2012, Yang et al. 2023). In agroecosystems, ET is the dominant consumptive water flux, with large implications for the soil water balance (McMillan 2020, Zaerpour et al. 2025), drought monitoring (Lu et al. 2025), irrigation scheduling (Wang et al. 2021), and, ultimately, crop production (Yang et al. 2021). Accurate quantification of ET is essential for improving agricultural water use and safeguarding regional water security and ecosystem integrity (Jha et al. 2019, Wanniarachchi and Sarukkalghe 2022). The influence of ET on regional water

balance is particularly important in semi-arid and Mediterranean climate zones, where crop water use is tightly constrained by the interaction between limited soil water availability and high atmospheric evaporative demand (Geerts and Raes 2009, Webber et al. 2025). In these rainfall-deficit environments, field hydrological processes are further modified by management, as cropping systems often alternate between severe soil water stress under water-limited conditions and well-watered conditions (Wu et al. 2021). As climate warming intensifies vapour pressure deficit (VPD), heat extremes, and within-season soil water depletion, pressure to expand irrigation in these regions is likely to increase (Trnka et al. 2014, Milly and Dunne 2016). Accurate estimation of ET in these cropping systems is a prerequisite for sustainable water management and climate-resilient agriculture.

To date, most assessments of simulated ET in cropping system models have considered performance metrics such as root mean square error (RMSE) and mean bias, which often produce inconsistent ranks of ET simulation methods across

sites and water-supply conditions (Gupta et al. 2026). Absolute metrics may also partly reflect the magnitude of observed ET, whereas mean bias alone does not diagnose mismatch in day-to-day dynamics. Previous studies (Lian et al. 2018, Webber et al. 2025) have indicated that daily ET simulations commonly exhibit systematic bias, while absolute errors often increase sharply with rising atmospheric evaporative demand, suggesting that ET simulation error can be influenced by multiple environmental factors. However, the main sources of error remain unclear. Simulation errors may arise from model structural limitations, daily climatic variability, or compensating errors whereby models offset weak sensitivity to atmospheric water demand (Chen et al. 2025), such as VPD, by overestimating crop sensitivity to soil water stress. Without separating systematic bias, mismatch in response magnitude, and lack of correlation, it is difficult to identify the physical basis of model failure under contrasting hydrological conditions (Webber et al. 2025). In turn, understanding the sources of model error is key to improving simulation of ET.

ET simulation error can vary strongly with environmental conditions (Qiu et al. 2025). Models that perform well under non-water-limited conditions often show substantial underestimation when soil water becomes limiting and atmospheric demand is high, such as under dry or windy conditions (Kang et al. 2009, Kimball et al. 2023, Su and Singh 2023). A recent global meta-analysis (Ding et al. 2025) further showed that incorporating soil moisture (SM) information can markedly improve the accuracy of crop ET estimation, suggesting that ET error is associated not only with meteorological forcing (Wu et al. 2021) but also with the nonlinear interplay between water-supply status and atmospheric demand. However, in the sensitivity analysis performed by Webber et al. (2025), atmospheric demand-related processes explained much more variation in daily ET simulations than did assumptions about root growth or soil water availability. Many simplified approaches neglect key processes, including heterogeneous root water uptake, nonlinear stomatal regulation and physiological lags during rewetting, and may therefore underestimate water-stress effects (Bruckler et al. 2004, Garrigues et al. 2015, Liu et al. 2020b).

Despite the strong influence of irrigation and soil moisture on ET dynamics, few studies (Ding et al. 2025, Webber et al. 2025) have explicitly examined how environmental associations with daily ET simulation error differ between water-limited and irrigated regimes. As a result, it remains unclear whether water-supply regime alters only absolute error magnitude, whether apparent differences persist after normalization, or whether it also changes error structure, ET-method dependence, and climatic associations. Under water-limited conditions, ET is typically constrained by progressive soil water depletion and plant hydraulic stress (Liu et al. 2020b, Vereecken et al. 2022). Under irrigation, by contrast, the partial relaxation of water limitation on crop water status can amplify model sensitivity to atmospheric demand, including aerodynamic forcing, as well as to the parameterization of canopy structure (Cui et al. 2021, Zhao et al. 2025).

Against this background, the present study aimed to disentangle the sources and environmental associations of daily

wheat ET simulation error, defined as the deviation of simulated ET from observed ET, in semi-arid and Mediterranean environments. Using simulations output from 17 wheat crop models, daily ET simulation error was quantified, and its error structure was characterized. We then examined whether error magnitude and structure varied among ET-demand approaches, developmental stages, and water-supply regimes. Mixed-effects models were used to assess the conditional associations of meteorological factors with daily ET error. Finally, focusing on water-limited conditions, we evaluated whether observed soil moisture and its interaction with atmospheric VPD provided additional information about daily ET simulation error.

2. Materials and methods

2.1 Experimental data and simulation data

Observational data were derived from field experiments conducted at two wheat-growing sites: Bushland, TX, USA and Avignon, France. The Bushland site, located in the semi-arid southern High Plains of the United States, provided paired water-supply treatments across three winter wheat growing seasons, including full irrigation and water-limited treatments (Evelt et al. 1994, Howell et al. 1995). The Avignon site contributed four predominantly water-limited durum wheat seasons, one of which received limited supplemental irrigation but remained classified as water-limited. For all analyses, treatments were assigned to either a fully irrigated or water-limited regime; the latter denotes rainfed or otherwise non-fully irrigated treatments and was not treated as a separate deficit-irrigation category. Available observations for each season included daily meteorological variables, soil profile properties, crop management information, phenological observations, repeated measurements of leaf area index (LAI), aboveground biomass and grain yield, as well as soil profile water content. At Bushland, daily ET was measured using weighing lysimeters, whereas at Avignon, daily ET was derived from an eddy-covariance system after rigorous quality control (Evelt et al. 2016). Because quality-control filtering removed a substantial number of observational days from the eddy-covariance record, the Avignon dataset contained discontinuous but high-confidence daily ET observations (Bruckler et al. 2004, Garrigues et al. 2015), while the Bushland dataset provided nearly continuous daily records. Soil water content was determined using neutron probes. Depending on the season, the maximum observation depth reached approximately 190–230 cm at Bushland and 150–190 cm at Avignon. A full description of the experimental datasets was provided by Webber et al. (2025).

The simulation protocol was described in detail by Webber et al. (2025), with the main aspects summarized here. The simulations were conducted as part of the AgMIP wheat model team activity (Rosenzweig et al. 2013). This ensemble comprised 17 distinct crop model structures contributed by 13 modelling groups (Webber et al. 2025). All participating models followed the intercomparison protocol and were calibrated using the full set of crop-growth observations available for each experiment, including phenology, LAI, biomass, and yield. Simulations were initialized on the

date for which initial soil–water conditions were provided, typically shortly before sowing, and were run until model-specific physiological maturity. In this study, daily ET outputs from these calibrated simulations were compared with the corresponding observed ET records, forming the basis for all subsequent error analyses.

Records with missing observed ET were excluded. To avoid unstable performance estimates from sparsely observed experiments, experimental trials with fewer than 30 valid ET-observation days were removed before analysis. All subsequent error metrics were calculated from matched daily simulated and observed ET records after this filtering.

To examine structural differences in ET simulation schemes, the 17 model outputs were classified into three groups according to their representation of atmospheric evaporative demand: reference crop evapotranspiration (ET_o), Priestley–Taylor potential evapotranspiration (PET-PT), and other potential evapotranspiration methods (PET-other). The ET_o group primarily used the FAO-56 Penman–Monteith formulation (Allen et al. 1998) combined with a crop coefficient. PET-other comprised the remaining potential-ET approaches not based on Priestley–Taylor. The classification was used to test whether ET simulation errors were associated with broad differences in evaporative-demand formulation rather than with individual model identities. The crop models and their ET-demand classifications are summarized in [Supplementary Table S1](#).

As the participating crop models differed in their representation and terminology of crop development, we applied a common three-stage phenological classification to enable consistent comparisons across models. The classification distinguished early vegetative growth, full canopy cover, and canopy senescence using LAI-based criteria established during data preparation. Early vegetative growth represented the pre-anthesis period with LAI < 0.5 m² m⁻², when soil evaporation could still contribute substantially to total ET. Full canopy cover comprised days with LAI ≥ 0.5 m² m⁻² before the onset of canopy decline, when transpiration was expected to dominate ET. Canopy senescence represented the period after maximum LAI had been reached and post-anthesis canopy decline had begun. The common phenological stage classification was identical across models within each trial and date, thereby providing a consistent phenological reference for the stage-specific analyses. Because the classification was derived from the ensemble median rather than from each model's own simulated LAI, the stage label was identical across models within each trial and date. Stage-specific comparisons among ET-demand classes are therefore not confounded by between-model differences in simulated LAI. To assess whether differences between the common and model-specific stage assignments affected the results, we repeated the stage-specific mean squared error (MSE) comparisons using only model days for which the two classifications agreed ([Supplementary Fig. S1](#); [Supplementary Table S7](#)).

2.2 Simulated ET error metrics and statistical analyses

Daily simulated ET error was quantified as ET_{diff},

$$ET_{diff} = ET_{sim} - ET_{obs} \quad (1)$$

where ET_{sim} indicates simulated ET and ET_{obs} indicates observed ET, both at the daily time step. Daily squared error was calculated as (ET_{sim} – ET_{obs})² and MSE was calculated within each model × experimental trial × water-supply regime unit to compare model performance among groups.

To distinguish absolute error magnitude from relative model performance, we calculated RMSE and scale-normalized error metrics for the same model × experimental trial × water-supply-regime units used in the MSE analysis. The principal normalized metric was RMSE divided by mean observed ET (NRMSE), while normalization by the standard deviation and range of observed ET, together with relative bias, was used to assess sensitivity to the choice of scaling method. Undefined ratios were treated as missing ([Supplementary Fig. S4](#); [Supplementary Table S6](#)).

To further diagnose the structure of ET simulation error, we decomposed MSE following [Gauch et al. \(2003\)](#) into squared bias (SB), non-unity slope (NU), and lack of correlation (LC). For a given paired simulated (X_i) and observed (Y_i) daily ET values, MSE can be divided as:

$$MSE = SB + NU + LC \quad (2)$$

where:

$$SB = (\bar{X} - \bar{Y})^2 \quad (3)$$

$$NU = (1 - b)^2 \times \frac{\sum_{i=1}^N (X_i - \bar{X})^2}{N} \quad (4)$$

$$LC = (1 - r^2) \times \frac{\sum_{i=1}^N (Y_i - \bar{Y})^2}{N} \quad (5)$$

Here, $\bar{X}$ and $\bar{Y}$ are the means of simulated and observed ET, b is the slope of the least-squares regression of observed ET on simulated ET, and r^2 is the squared correlation coefficient between simulated and observed ET. In this framework, SB quantifies the systematic component associated with persistent over- or underestimation, NU represents deviation from a unity slope and therefore mismatch in response magnitude, and LC reflects the remaining error associated with imperfect correlation between simulations and observations. This decomposition was selected because it provides distinct additive components with direct interpretations of prediction–error structure.

The decomposition was applied to compare error structure across water-supply regimes, sites, developmental stages, and ET-method classes. Group differences in MSE and in the relative contributions of SB, NU, and LC were evaluated using non-parametric tests because the distributions were non-Gaussian ([Hollander et al. 2013](#)). Overall contrasts were tested using Kruskal–Wallis or Wilcoxon rank-sum tests, depending on the number of groups, and pairwise Wilcoxon rank-sum test P values were adjusted within each comparison family using the Benjamini–Hochberg false discovery rate procedure. The same aggregation unit and corresponding rank-based comparison were used for the normalized error metrics to maintain comparability with the original MSE analysis ([Supplementary Table S6](#)).

2.3 Assessing climatic sources of simulated ET error

To attribute daily simulated ET error to meteorological factors, a set of daily climatic predictors was assembled for each observation record, including maximum air temperature ($T_{\max}$), diurnal temperature range (DTR), wind speed, shortwave radiation, precipitation, and VPD. Saturation vapour pressure (e_s) was calculated using the following formulation (Allen et al. 1998), which is a core formulation for FAO-56 method:

$$e_s(T) = 0.6108 \times \exp\left(\frac{17.27 T}{T + 237.3}\right) \quad (6)$$

with $e_s(T_{\max})$ and $e_s(T_{\min})$ averaged to obtain $e_{s,mean}$. T represents the temperature. VPD was computed as follows:

$$VPD = \max(e_{s,mean} - e_a, 0) \quad (7)$$

Because VPD is strongly coupled with air temperature, its collinearity with mean temperature (T_{mean}) and maximum temperature ($T_{\max}$) was examined. The actual vapour pressure was obtained from measured vapour-pressure information where available or estimated from measured relative humidity at the daily maximum and minimum temperatures.

$T_{\max}$ showed the stronger association with VPD and was therefore used to derive a VPD-independent temperature term. Within each water-supply regime, $T_{\max}$ was regressed against VPD and the residuals were retained:

$$T_{\max} = \alpha_0 + \alpha_1 VPD + \varepsilon \quad (8)$$

In the linear regression equation, α_0 is the intercept while α_1 is the regression slope, representing the average change in $T_{\max}$ associated with a one-unit increase in VPD. The residual term, ε , hereafter denoted as $T_{\max,resid}$, represents the component of $T_{\max}$ not linearly explained by VPD. It was used to test whether a VPD-independent heat signal improved attribution of simulated ET errors while reducing collinearity in the mixed-effects models.

All climatic attribution analyses were conducted separately for water-limited and irrigated subsets. Linear mixed-effects models were fitted with daily simulated ET error as the response variable, climatic factors as fixed effects, and experimental repetition (denoted as trial in Equations 9 and 10) as a random intercept to account for repeated daily observations within the same experiment. Candidate fixed effects included VPD, residual maximum temperature ($T_{\max,resid}$), wind speed, DTR, shortwave radiation, and precipitation together with ET-demand approach class and developmental stage as categorical covariates. Models were fitted by maximum likelihood when comparing alternative fixed-effects structures. Model identity was not included as an additional random intercept because model structures were largely nested within ET-demand approach class, so that between-model heterogeneity could not be separated cleanly from the fixed ET-demand class effect in the available dataset. Trial was therefore retained as the sole random intercept to account for repeated daily observations within experiments.

To evaluate whether the VPD-independent heat term ($T_{\max,resid}$) improved model fit beyond VPD alone, two candidate structures were compared within each water-supply

regime: a VPD-only model and a model including both VPD and $T_{\max,resid}$ (Equations 9 and 10). Model comparison used the Akaike information criterion (AIC), Bayesian information criterion (BIC), and marginal and conditional explained variance. Marginal R^2 represents the variance explained by the fixed effects alone, whereas conditional R^2 represents the variance explained by both fixed and random effects. After selecting the preferred model structure within each water-supply regime, drop-one analyses were conducted by removing each climatic predictor individually while retaining all other terms in the model. Joint removal of multiple predictors was not considered. For each predictor, ΔAIC was calculated as $AIC_{reduced} - AIC_{full}$, where the reduced model excluded the focal predictor and the full model retained all terms. Larger positive ΔAIC values therefore indicate a greater loss of model support following removal of that predictor. Because each predictor was evaluated conditional on the remaining covariates, these values represent conditional contributions to model fit and should not be interpreted as causal effects.

Fixed-effect coefficients and 95% confidence intervals were examined to quantify the direction and magnitude of each meteorological association with ET_{diff} conditional on the other covariates.

$$ET_{diff} \sim \alpha_0 + VPD + \dots + (1 | \text{trial}) \quad (9)$$

$$ET_{diff} \sim \alpha_0 + VPD + T_{\max,resid} + \dots + (1 | \text{trial}) \quad (10)$$

To interpret categorical effects, estimated marginal means were computed for phenological stage and ET-method class within the selected mixed-effects models. Marginal response curves were generated by varying one focal predictor across its observed range while holding the remaining predictors at representative values, thereby visualizing the partial response of ET_{diff} to individual climatic factors. Climatic associations with daily error magnitude were additionally examined using log-transformed squared error (logSE, defined as $\log((ET_{diff})^2)$). This formulation was used to assess whether the effects of these factors on ET_{diff} were consistent with those on error magnitude. The same predictors and mixed-effects modelling framework were applied as in the main analysis.

Because the main models used daily records without an explicit residual correlation structure, temporal dependence was assessed by summarizing short-lag residual autocorrelation within repeated-observation groups and by refitting the selected models to non-overlapping 3-day aggregated records. This sensitivity analysis was used to evaluate the directional robustness of fixed-effect estimates rather than to establish independence of the daily residuals (Supplementary Fig. S3; Supplementary Table S9).

2.4 Assessing soil moisture impact on ET simulation error

Because soil–water availability mediates the translation of atmospheric demand into ET error, additional analyses quantified the role of observed SM in explaining ET_{diff} . A dataset linking ET observations to SM measurements was constructed because daily ET observations and SM measurements were not

always recorded on the same dates. The analysis was restricted to water-limited experiments, in which soil–water availability was expected to exert the strongest constraint on ET, while matched intervals containing recorded wetting events were excluded. In the absence of recorded precipitation, SM was assumed to remain sufficiently stable over the short matching interval for the matched measurement to represent the ET observation date.

Matching rules were adapted to site-specific data availability, using a 5-day matching window at both sites. At Avignon, where soil–moisture observations were relatively frequent, ET observation days were used as anchors. For each ET observation, the nearest preceding soil–moisture measurement within a 5-day backward window was selected. At Bushland, where soil–moisture observations were less frequent, soil–moisture measurement days were used as anchors to maximize sample retention. For each soil–moisture observation, the earliest subsequent ET observation within a 5-day forward window was selected. At both sites, matched pairs were retained only when no wetting event was recorded within the corresponding search window, based on the available daily precipitation and irrigation records. This procedure reduced the likelihood of assigning soil–moisture conditions that had been altered by recent rainfall or irrigation. The final matched dataset linked each retained ET observation day, together with the corresponding simulated ET values from each ET-demand approach, to a common antecedent soil–moisture state. Site-specific windows, gap distributions, SM-date reuse, and date- versus method-level sample sizes are reported in [Supplementary Table S8](#).

Based on this dataset, correlation coefficients were calculated between daily simulated ET error (ET_{diff}) and SM measured at each available depth, and then separately between absolute simulated ET error ($|ET_{diff}|$) and soil moisture, within each ET-demand approach class. To facilitate analysis and attribution, measurement depths were aggregated into three functional layers representing distinct hydrological controls: shallow soil moisture, root-zone soil moisture, and deep soil moisture. The shallow layer captured near-surface wetness (defined as the mean volumetric soil water content across 10–30 cm depth), the root-zone layer represented soil moisture conditions in the main rooting domain (calculated as the mean volumetric soil water content across 50–90 cm depth), and the deep layer represented deeper subsurface storage as the mean volumetric soil water content below 1 m. Baseline models without soil–water terms were compared with models including each SM layer as an additive predictor. Improvement in fit was quantified as the reduction in AIC relative to the corresponding baseline model.

Finally, to test whether atmospheric demand and subsurface water supply jointly shaped ET errors under water-limited conditions, interactions between VPD and soil moisture were evaluated using mixed-effects models. Separate candidate models were constructed for shallow and root-zone soil moisture. The added value of each interaction was assessed by comparing AIC with the corresponding additive model, and the interaction coefficient was used to determine whether the association between atmospheric demand and ET_{diff} changed with soil–water availability. Predicted ET_{diff}

surfaces across the observed VPD–soil–moisture space was generated from the selected interaction model. To evaluate sensitivity to reuse of the same SM measurement across multiple matched records, the SM analyses were repeated after aggregation to the SM-date level and, separately, after retaining one ET date per SM date; the selection rule, sample-size changes, and resulting model comparisons are documented in [Supplementary Fig. S2](#) and [Supplementary Table S8](#). All analyses were performed in R version 4.3.3. Data handling and visualization used *tidyverse* packages, including *dplyr*, *tidyr*, and *ggplot2*; mixed-effects models used *lme4* and *lmerTest*; and model evaluation and output extraction used *performance* and *broom.mixed*.

3. Results

3.1 Determination of ET simulation errors

Our comparison of absolute ET error showed that RMSE was significantly higher under irrigated than under water-limited conditions ([Fig. 1A](#), $P < .001$). This contrast was reflected in both a higher median RMSE and a broader distribution under irrigation. However, Avignon contributed only water-limited experiments, whereas Bushland contributed both water-limited and fully irrigated experiments; the pooled contrast therefore describes the available datasets and should not be interpreted as an isolated causal effect of irrigation.

After normalization by mean observed ET, the contrast between water-supply regimes was substantially attenuated ([Fig. 1B](#)). NRMSE remained slightly higher under irrigation, but the difference did not reach the conventional .05 significance level ($P = .073$). Relative bias also did not differ clearly between regimes ($P = .327$), and alternative normalizations by the standard deviation and range of observed ET did not indicate consistently poorer relative model performance under irrigation ([Supplementary Fig. S4](#); [Supplementary Table S6](#)). Thus, irrigation was associated with larger absolute ET errors, but the available evidence did not demonstrate uniformly poorer relative performance.

Decomposition of total MSE further showed that the dominant source of ET error was similar in type across regimes, but differed in magnitude ([Fig. 1C](#)). Under both water-limited and irrigated conditions, lack of correlation (LC) was the largest component, accounting for 64.1% and 74.5% of total MSE, respectively. By contrast, SB contributed 24.8% under water-limited conditions and 17.0% under irrigation, while the NU slope component contributed 11.1% and 8.5%, respectively. These results indicate that absolute ET error arose primarily from imperfect correspondence with the observed day-to-day temporal pattern rather than from persistent mean bias. Given the attenuation of the regime contrast after normalization, however, the larger absolute error and LC contribution under irrigation should not be interpreted as evidence of uniformly poorer relative model performance.

3.2 Model structural and phenological sources of error

Across the full dataset, daily ET simulation error differed among the broad ET-demand approach classes implemented

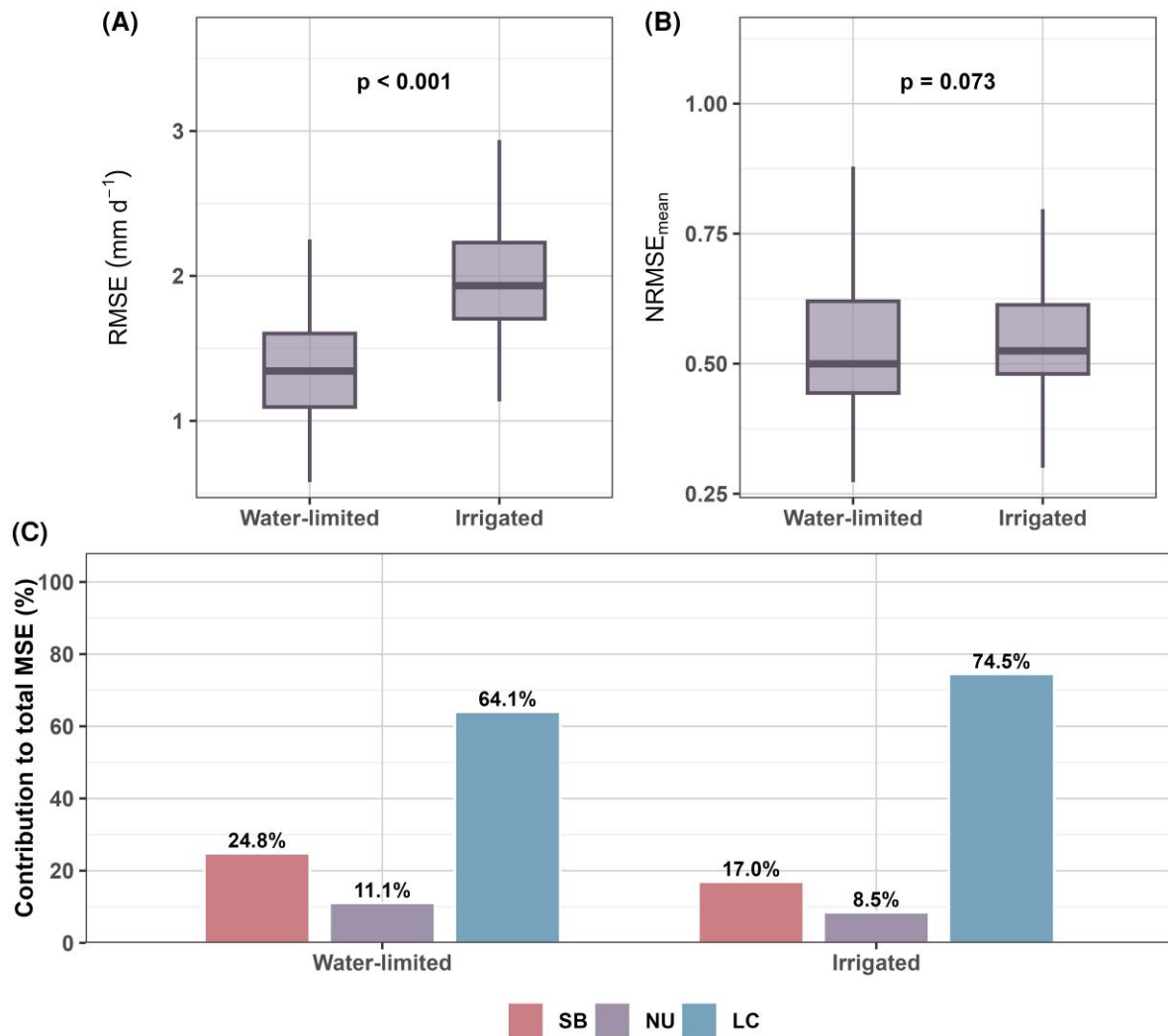


Figure 1 Water-supply dependence of absolute and normalized ET simulation error and its structural decomposition. (A) Distribution of RMSE under water-limited and irrigated conditions. (B) RMSE normalized by mean observed ET (NRMSE_{mean}) under each water-supply regime. Avignon contributed only water-limited experiments, whereas Bushland contributed both water-limited and irrigated experiments. The pooled comparison therefore describes the available datasets and should not be interpreted as an experimentally isolated effect of irrigation. (C) Percentage contributions of SB, NU, and LC to total MSE under each water-supply regime. In the boxplots, the centre line denotes the median, boxes indicate the interquartile range, whiskers extend to $1.5 \times \text{IQR}$, and points represent model-by-trial units.

in the crop models (Kruskal–Wallis, $P = .03$; Fig. 2A). This overall contrast was driven primarily by greater MSE under PET-PT than under ETo ($P = .03$), whereas ETo and PET-other ($P = .54$) and PET-PT and PET-other ($P = .15$) had similar errors. Thus, although the three classes showed broadly overlapping error distributions, PET-PT tended to have higher absolute ET error than ETo in the pooled analysis. Because models within each class also differed in other structural components and calibration choices, these contrasts cannot be attributed exclusively to the ET-demand formulation.

The error dependence of the ET-demand simulation approach varied across wheat growth stages. During the early vegetative stage, differences in MSE among ET-demand approaches had already emerged (Fig. 2B), with PET-other showing the highest median error and PET-PT the lowest. The difference in MSE among ET-demand approaches increased during full canopy cover (Kruskal–Wallis, $P = .04$; Fig. 2C),

with PET-PT showing the highest median MSE. By contrast, during senescence, differences among ET-demand approaches were no longer evident (Kruskal–Wallis, $P = .24$; Fig. 2D), and all classes showed wider error dispersion and several extreme values.

Restricting the analysis to model-days for which the common and model-specific stage assignments agreed broadly retained these stage-dependent patterns (Supplementary Fig. S1; Supplementary Table S7). The early-stage ordering remained PET-PT < ETo < PET-other, while the full-cover ordering ETo < PET-other < PET-PT was also retained, although the overall full-cover contrast was weaker but still significant at 0.1 level in the aligned subset. Differences during senescence remained weak ($P = .16$), and the relative ordering of PET-PT and PET-other was less stable. Thus, the broad stage-specific patterns were retained, but method separation during full cover and especially senescence was sensitive to stage alignment.

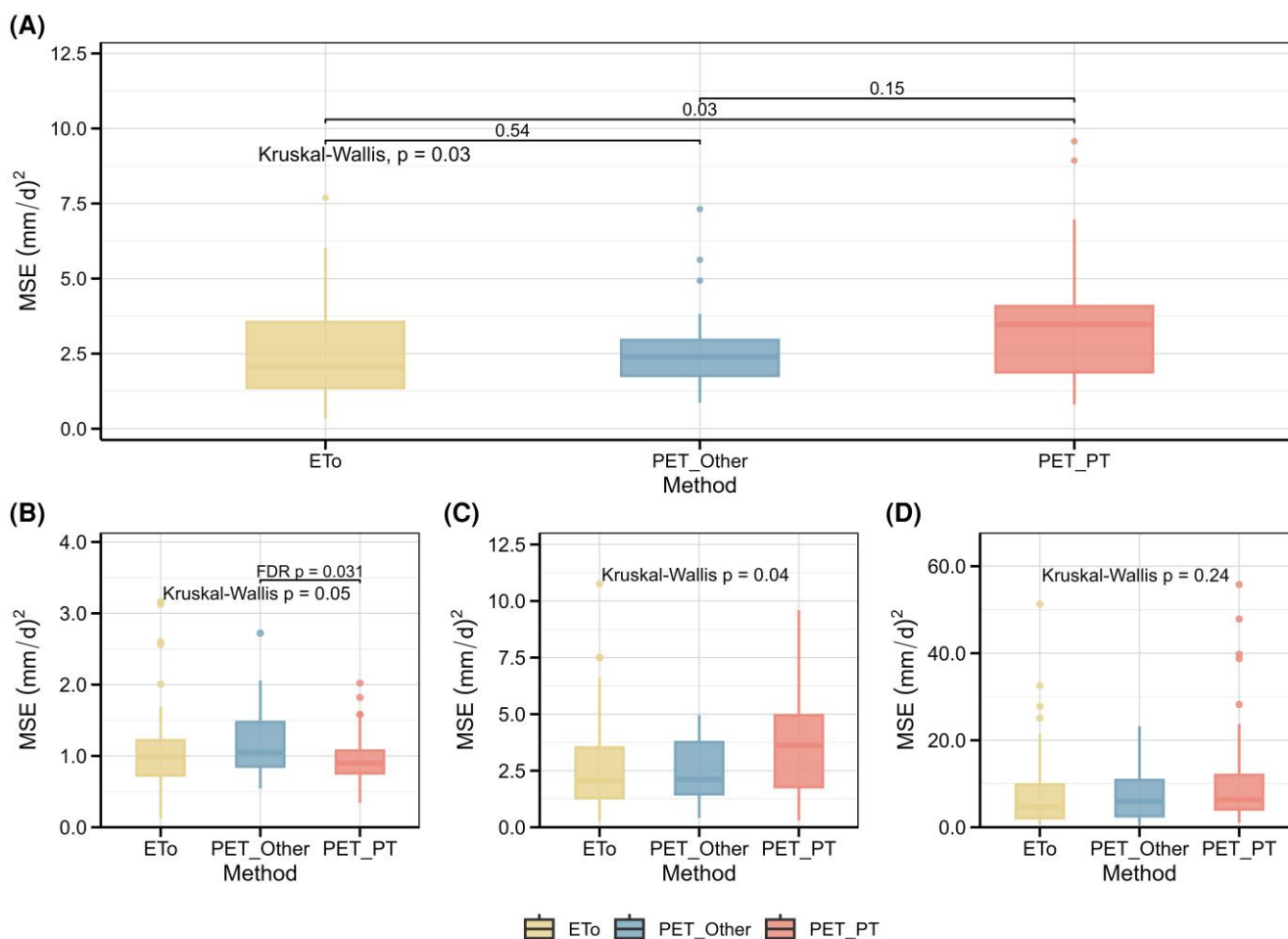


Figure 2 Method dependence of ET simulation error across both sites and all crop development stages. Boxplots show the distribution of MSE of daily ET simulations for the three ET-demand simulation approaches (ET_o, PET-other, and PET-PT). (A) all observations pooled across sites, water-supply regimes, and development stages. (B–D), Results stratified by crop development stage: (B), early vegetative stage; (C), full-cover stage; (D), senescence stage. Numbers above brackets indicate pairwise Wilcoxon rank-sum test *P* values adjusted within each developmental stage using the Benjamini–Hochberg false discovery rate procedure; inset text reports the unadjusted overall Kruskal–Wallis test. Boxes represent the interquartile range, centre lines indicate medians, whiskers extend to 1.5 × the interquartile range, and points denote outliers.

As shown in the previous section, absolute ET error was higher under irrigation in the datasets available here, although the contrast was attenuated after normalization. Further examination assessed whether errors associated with the different ET-demand approaches varied between water-supply regimes (Fig. 3). Under water-limited conditions, daily ET simulation error was similar among ET-demand approaches (Kruskal–Wallis, $P = .341$; Fig. 3A), and none of the pairwise comparisons was significant ($P = .31$, $.20$, and $.84$). Under fully irrigated conditions, PET-PT tended to have higher MSE than ETo and PET-other, but the overall test did not reach the conventional $.05$ significance level (Kruskal–Wallis, $P = .078$; Fig. 3B). The PET-PT versus PET-other contrast reached significance ($P = .041$), whereas the PET-PT versus ETo contrast was marginal ($P = .058$), and ETo and PET-other remained similar ($P = .75$). Notably, the median errors of PET-PT and ETo were relatively similar, but ETo showed greater within-group variability, including a subset of simulations with errors close to zero, which increased overlap between the two distributions.

3.3 Climatic determinants of simulated ET error

To assess how climatic factors were associated with daily ET simulation error, we first compared two mixed-effects models within each water-supply regime: a VPD-only model and an extended model that added $T_{\text{max, resid}}$, the component of maximum temperature not linearly explained by VPD. Because VPD was strongly coupled with air temperature and showed a tighter association with T_{max} than with T_{mean} (Supplementary Table S2), this comparison was used to assess whether daytime maximum temperature provided additional information beyond VPD and to establish the baseline model for subsequent analyses (Fig. 4A, B).

The extended model improved fit within both water-supply regimes. Under water-limited conditions, adding $T_{\text{max, resid}}$ reduced AIC from 6833.6 to 6764.7 and BIC from 6903.4 to 6840.3, while marginal and conditional R^2 increased from 0.16 to 0.18 and from 0.26 to 0.30, respectively. Under irrigated conditions, AIC declined from 6740.0 to 6545.2 and BIC from 6808.1 to 6619.0, while marginal and conditional R^2 increased from 0.40 to 0.45 and from 0.41 to 0.47,

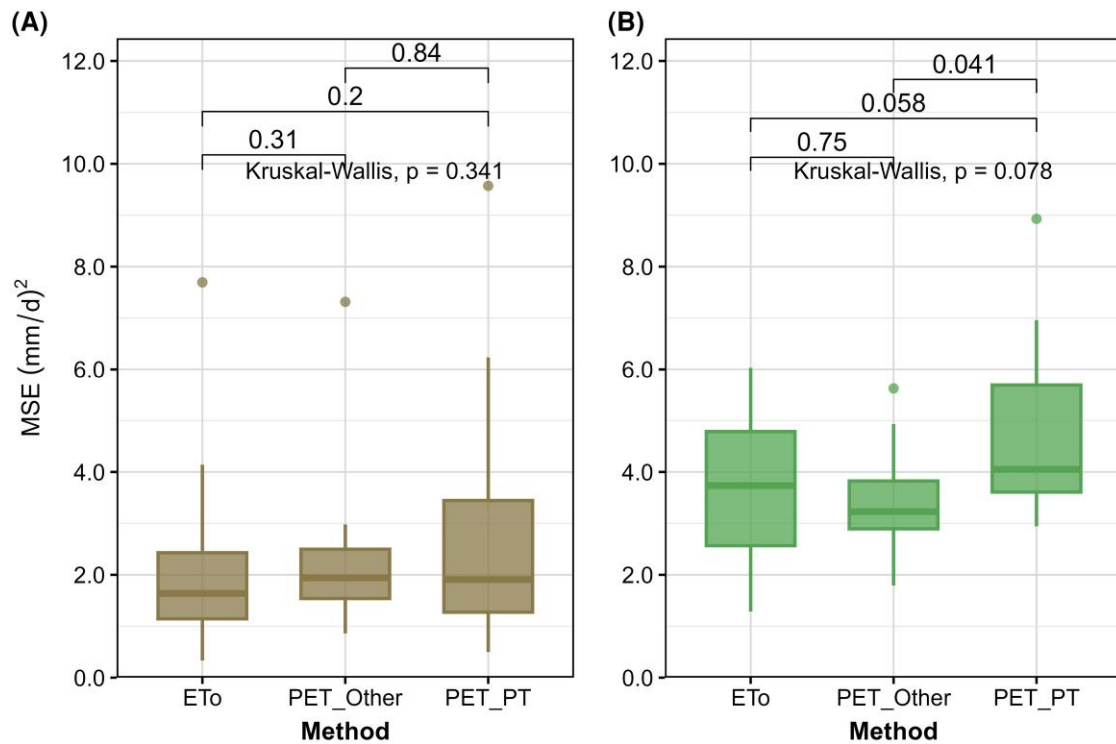


Figure 3 Water-supply dependence of the ET-demand approach effects on daily ET simulation error. Boxplots show the distribution of MSE among the three ET formulations (ET_o, PET-other, and PET-PT) under (A) water-limited or (B) fully irrigated conditions. Boxes indicate the interquartile range, centre lines denote medians, whiskers extend to 1.5 times the interquartile range, and points represent outliers. *P* values from Kruskal–Wallis tests are shown within each panel, and brackets indicate pairwise *post hoc* comparisons between ET formulations.

respectively. Because the regime-specific models were fitted to different observations, the absolute magnitudes of these information-criterion changes were not used to compare the strength of the temperature association between regimes. The extended model was retained as the baseline model within each regime.

Based on the baseline model, we quantified the conditional contribution of individual meteorological predictors using a drop-one analysis within each water-supply regime (Fig. 5). Under irrigated conditions, the largest deterioration in model fit occurred when $T_{\max, \text{resid}}$ was removed ($\Delta\text{AIC} = 194.8$), followed by wind speed ($\Delta\text{AIC} = 172.5$), radiation ($\Delta\text{AIC} = 95.1$), and VPD ($\Delta\text{AIC} = 71.7$). Precipitation made a comparatively smaller contribution ($\Delta\text{AIC} = 18.6$), whereas DTR provided little additional information. Under water-limited conditions, removing wind speed caused the largest deterioration ($\Delta\text{AIC} = 120.5$), followed by radiation ($\Delta\text{AIC} = 101.6$) and $T_{\max, \text{resid}}$ ($\Delta\text{AIC} = 69.0$). Precipitation contributed comparatively little ($\Delta\text{AIC} = 8.7$), while VPD and DTR added little conditional support after accounting for the other covariates.

The drop-one analysis therefore indicated that the conditional structure of signed ET error differed between the two water-supply datasets. These ΔAIC values describe changes in model support conditional on the fitted covariate structure and should not be interpreted as causal contributions or as directly comparable measures of predictor importance across the regime-specific datasets.

The fixed-effect coefficients of the selected attribution model were then examined to quantify the direction and

magnitude of each estimated association with daily ET error, conditional on the other climatic covariates and accounting for trial-specific random intercepts (Table 1). The coefficient patterns differed between water-supply regimes, indicating that the fitted associations between ET error and climatic factors varied between the water-limited and fully irrigated datasets.

Under fully irrigated conditions, ET_{diff} was strongly associated with both the VPD and VPD-independent temperature component. $T_{\max, \text{resid}}$ showed a positive association with ET_{diff} , whereas VPD and wind speed were both negatively associated with ET_{diff} , with narrow confidence intervals and strong statistical support. Shortwave radiation also exhibited a negative relationship, while precipitation was positively associated with ET_{diff} . By contrast, DTR showed no evident independent effect once the other predictors were included.

Under water-limited conditions, wind speed and radiation remained negatively associated with ET_{diff} , and $T_{\max, \text{resid}}$ retained a positive association. The VPD coefficient was weak and its confidence interval included zero (estimate = -0.098 , $P = .109$). Precipitation remained positively associated with ET_{diff} but still weak, whereas DTR again contributed little independent signal. The small conditional contribution of VPD in this model does not imply that VPD was physiologically unimportant; rather, it indicates that VPD added little information after the other correlated climatic and categorical predictors had been included.

The climatic effects were also examined using the simpler VPD-only model, which further indicated how inclusion of

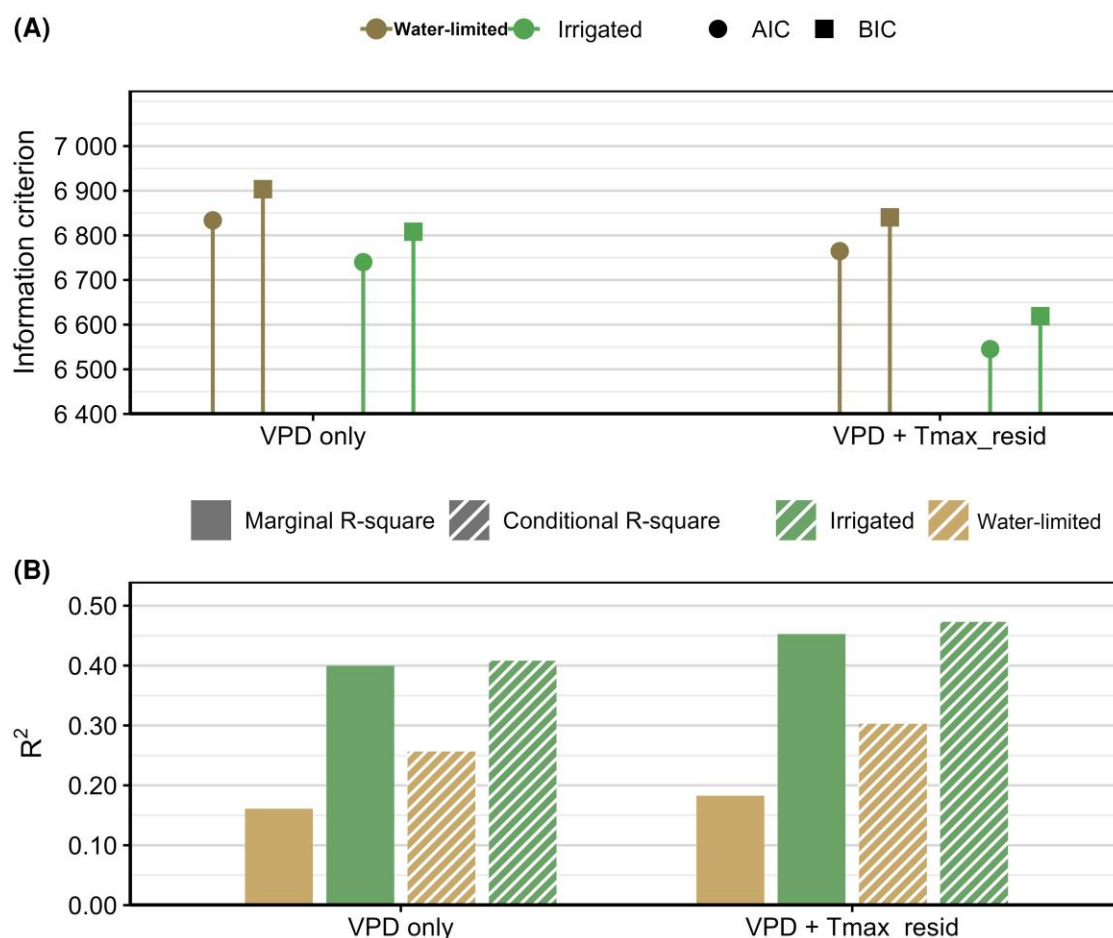


Figure 4 Comparison of mixed-effects attribution models for daily ET error under water-limited and irrigated conditions. (A) Model comparison based on AIC and BIC for two mixed-effects structures fitted separately under water-limited and irrigated conditions: the VPD-only model and an extended model including $T_{\max,\text{resid}}$, defined as the component of maximum temperature not linearly explained by VPD. Smaller AIC and BIC indicate stronger model support within each regime-specific dataset. (B) Marginal and conditional R^2 for the same models. Marginal R^2 represents variance explained by the fixed effects alone, whereas conditional R^2 represents variance explained by both fixed and random effects. Inclusion of $T_{\max,\text{resid}}$ improved model support within both regimes; because the models were fitted to different data subsets, the magnitudes of the information-criterion changes were not compared across regimes.

$T_{\max,\text{resid}}$ altered the coefficient patterns (Supplementary Table S3). In the simpler model, VPD showed a stronger negative association with ET_{diff} under both water-supply regimes. After inclusion of $T_{\max,\text{resid}}$, this negative VPD association was attenuated and, under water-limited conditions, became weak and non-significant. By contrast, $T_{\max,\text{resid}}$ emerged as a positive predictor in the extended model, indicating that part of the signal initially assigned to VPD was shared with the component of daytime maximum temperature not linearly explained by VPD. The coefficients for wind speed, radiation, and precipitation were broadly stable across the two model structures.

Apart from the climatic terms, the models also revealed systematic shifts associated with ET-demand approach and crop growth stage. Relative to the PET-other group, both ETo and PET-PT were associated with more negative ET_{diff} , with the stronger negative offset observed for PET-PT. These method-related differences were more pronounced under irrigation but should be interpreted as associations with groups of models rather than effects attributable solely to the ET-demand

equation. Relative to full cover, early vegetative conditions were associated with more positive ET_{diff} under irrigation, with weaker evidence under water-limited conditions. Senescence showed contrasting associations between both water supply regimes, with more positive ET_{diff} under water-limited conditions and more negative ET_{diff} under irrigation.

Residual diagnostics showed positive short-lag temporal autocorrelation in the daily mixed-effects models (Supplementary Fig. S3A; Supplementary Table S9), with median lag-1 autocorrelation of 0.512 under water-limited conditions and 0.354 under irrigation. After aggregation into non-overlapping three-observation bins, 19 of 20 non-intercept fixed-effect directions were retained (Supplementary Fig. S3B). The only reversal was the weak water-limited VPD coefficient, which was not statistically distinct from zero in either analysis. Most climatic, method, and stage associations were therefore directionally stable, although inference based on individual daily-scale P values should be interpreted cautiously.

A complementary analysis using log-transformed squared error (logSE) revealed partly different coefficient patterns

from the signed-error model (Supplementary Table S4). Under both water-supply regimes, higher VPD, wind speed, radiation, and precipitation were generally associated with larger error magnitude, whereas the VPD-independent temperature component showed a weak negative association under water-

limited conditions and a stronger negative association under full irrigation. These contrasts indicate that some climatic factors were primarily associated with the direction of daily simulated ET error, whereas others more consistently amplified its overall magnitude. PET-PT also tended to be associated with larger error magnitude than the reference ET method, while phenological stage effects were more evident under irrigation than under water-limited conditions.

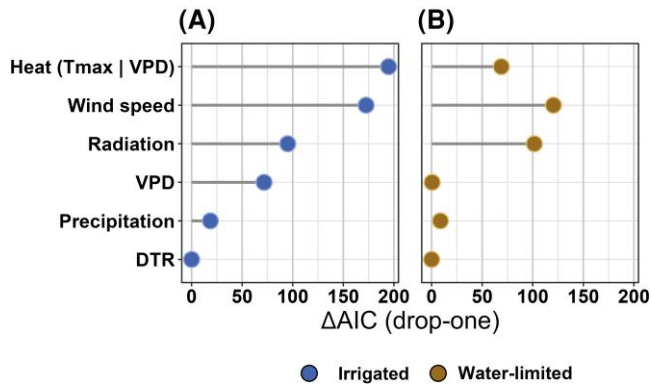


Figure 5 Conditional contributions of meteorological predictors to daily ET error under irrigated and water-limited conditions. Contributions were quantified using a drop-one procedure applied to the selected mixed-effects attribution model within each water-supply regime. For each predictor, the model was refitted after removing that term, and the deterioration in model fit was expressed as Δ AIC relative to the full model. Larger Δ AIC values indicate greater conditional contribution to model support within the same fitted model, but do not establish causal importance. Panel A shows irrigated conditions, and panel B shows water-limited conditions.

3.4 Influence of soil moisture content on ET simulation error

To examine whether soil–water status helped explain ET simulation errors under water-limited conditions, the relationship between observed soil moisture and ET_{diff} was analysed across the available measurement depths. The final matched dataset comprised 99 date-level ET–SM pairs, corresponding to 297 method-level records because each pair was represented for the three ET-demand approach classes (Supplementary Table S8). Soil moisture was generally positively correlated with signed ET_{diff} and negatively correlated with absolute ET_{diff} (Fig. 6). Thus, drier conditions tended to be associated with more negative and larger-magnitude ET errors, whereas wetter conditions were associated with errors closer to zero or more positive values. This pattern was strongest across both shallow and root-zone layers, with weaker correlations at greater depths, although PET-PT remained significantly related to deep soil moisture up to 150 cm. In contrast, the absolute daily simulated ET (Fig. 6b) error was negatively correlated with soil moisture across nearly all depths and methods, indicating that ET simulations were less accurate

Table 1 Fixed-effect estimates of the mixed-effects attribution model for daily simulated ET error (ET_{diff}) under water-limited and fully irrigated conditions.

Water supply	Variable type	Term	Estimate	95% CI low	95% CI high	Significance
Water-limited	Meteorological factors	Intercept	0.82	0.40	1.24	<.05
		VPD	−0.098	−0.22	0.02	.1086
		T_{max_resid}	0.048	0.04	0.06	<.05
		Wind speed	−0.14	−0.17	−0.12	<.05
		DTR	−0.0009	−0.01	0.01	\
		Radiation	−0.050	−0.06	−0.04	<.05
	ET method	Precipitation	0.022	0.01	0.04	<.05
		ET_o	−0.11	−0.20	−0.02	<.1
		PET-PT	−0.16	−0.25	−0.07	<.05
	Phenological stages	Early. veg	0.091	−0.01	0.19	<.1
		Senescence	0.34	0.20	0.49	<.05
Full irrigation	Meteorological factors	Intercept	1.90	1.49	2.32	<.05
		VPD	−0.56	−0.69	−0.44	<.05
		T_{max_resid}	0.098	0.08	0.11	<.05
		Wind speed	−0.22	−0.25	−0.19	<.05
		DTR	0.0040	−0.01	0.02	\
		Radiation	−0.059	−0.07	−0.05	<.05
	ET method	Precipitation	0.029	0.02	0.04	<.05
		ET_o	−0.23	−0.34	−0.11	<.05
		PET-PT	−0.45	−0.57	−0.34	<.05
	Phenological stages	Early.veg	0.17	0.05	0.29	<.05
		Senescence	−0.53	−0.71	−0.36	<.05

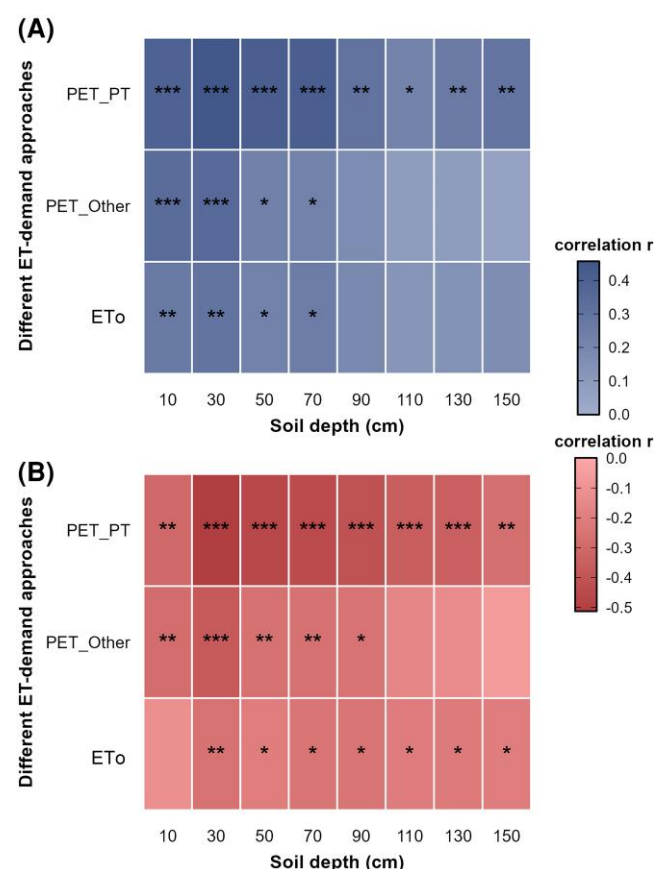


Figure 6 Correlations between soil moisture at different depths and daily simulated ET error (ET_{diff}) across ET methods. (A) Pearson correlation coefficients between SM and daily simulated ET error for different ET methods and soil depths. Positive values indicate that wetter soils are associated with more positive ET errors. (B) Pearson correlation coefficients between SM and the absolute ET simulation error. Negative values indicate that wetter soils are associated with smaller ET error magnitude. Asterisks denote statistical significance: $P < 0.05$ (*), $P < 0.01$ (**), and $P < 0.001$ (***).

under drier soil conditions and more stable under wetter soil conditions. These negative correlations were strongest for PET-PT and were most pronounced in the root-zone layers, suggesting that plant-available water, rather than surface wetness alone, may be an associated factor with ET model performance. These results therefore indicate that soil moisture was associated with both the direction and magnitude of ET simulation error, with root-zone moisture emerging as the most important index. Soil water was then grouped by depth into shallow (10–30 cm), root-zone (50–90 cm), and deep (over 1 m) layers in subsequent analyses because these layers represent different hydrological and physiological controls on ET.

Further analyses examined whether the daily simulated ET errors were directly associated with the daily soil-moisture simulation error. Across most soil layers and ET-demand approaches, correlations between ET and SM simulation errors were very weak and statistically non-significant (Supplementary Table S5). Only the PET-other approach showed significant relationships in the shallow and root-zone

layers, but these associations were modest and changed sign across depths. Overall, these results suggest that daily simulated ET errors may not be simply a direct propagation of coincident SM simulation errors. The subsequent analyses therefore focused on observed soil moisture as a hydrological state variable affecting daily simulated ET error, rather than on SM simulation error itself.

To quantify the contribution of SM to daily simulated ET errors, the selected climatic factors model was compared for a water-limited environment from our previous analysis—namely, the model shown in Equation 10—as the baseline model. The baseline model was then compared with corresponding models that additionally included SM terms at different depths and SM effects were evaluated among the three ET-demand approaches (Fig. 7A, B).

Overall, adding observed SM improved model fit relative to the climatic-only baseline, but the magnitude of improvement differed by depth. Among the single-factor SM models, root-zone SM provided the largest improvement, reducing AIC the most compared with the baseline model, whereas shallow SM produced a smaller improvement and deep SM added comparatively little support (Fig. 7A). These comparisons indicate that daily ET error was more closely associated with water availability in the principal rooting zone than with surface wetness or deeper subsurface storage alone.

The inset of Fig. 7A shows the drop-one analysis for the overall model and further supports a larger conditional contribution from root-zone SM. The *F*-values ranked as root-zone SM > shallow SM > deep SM, indicating that omission of root-zone SM caused the largest reduction in explanatory power among the tested layers. The association between observed SM and ET_{diff} also differed among ET-demand approaches (Fig. 7A, B). The root-zone SM × ET-demand approach term produced the largest improvement in model fit, followed by the shallow SM interaction, whereas the deep-SM interaction added little support. Method-specific comparisons showed that adding shallow and root-zone SM produced the largest AIC reductions for PET-PT, more moderate reductions for ETo, and smaller improvements for PET-other. These patterns indicate that observed SM was associated with daily simulated ET error to different degrees across ET-demand approach classes. Because the classes also differ in canopy, soil–water balance, and stress parameterization, these contrasts describe differences among broad model groups and do not isolate the representation of soil water itself.

VPD interactions with soil–water status were further investigated by comparing the baseline model with models including the main effect of SM and the VPD × SM interaction (Fig. 8A, B). Adding root-zone SM improved model fit relative to the baseline (AIC: 741.77–719.61), and inclusion of the VPD × root-zone SM interaction further reduced AIC to 708.67. In contrast, the VPD × shallow SM interaction produced almost no additional gain beyond the shallow-SM main-effect model (AIC: 730.30–729.93).

The negative interaction indicates that the association between VPD and daily simulated ET error depended on root-zone soil moisture. Under drier root-zone conditions, increasing VPD was associated with a stronger positive shift in ET_{diff}, whereas this association was weaker under wetter root-zone

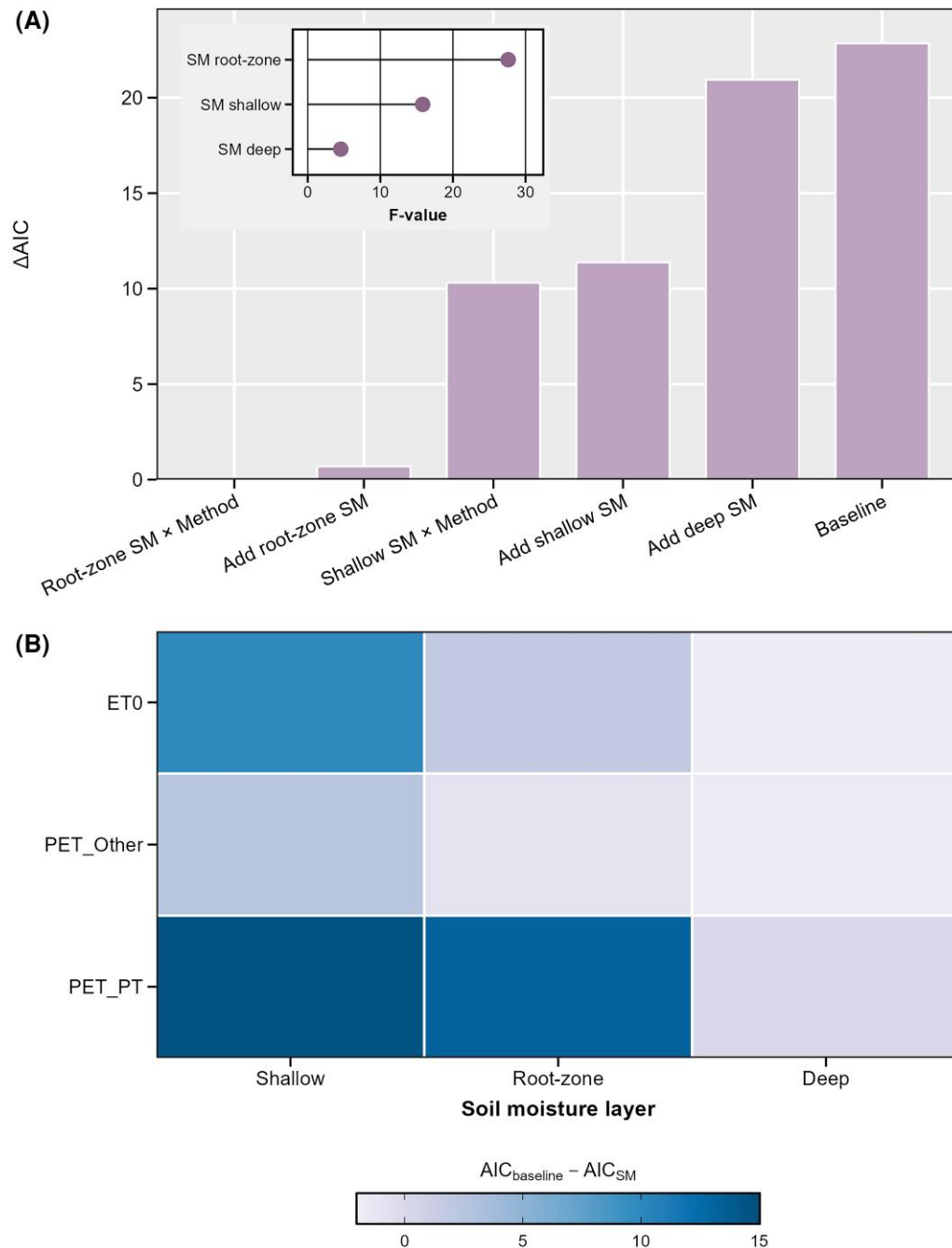


Figure 7 Overall and method-specific contributions of soil moisture to ET model performance. (A) Model comparison based on ΔAIC for models including SM at different depths and SM $\times$ method interaction terms. ΔAIC values are shown relative to the best-performing model, such that lower values indicate better model performance. The inset shows F-values from drop-one tests on the overall mixed-effects model, obtained by removing each SM layer term in turn while retaining the remaining predictors. Larger F-values indicate a greater loss of explanatory power when that SM term is omitted, and thus a greater independent contribution of that SM layer to explaining ET simulation error. (B) Method-specific heatmap of model improvement, expressed as $(AIC_{baseline} - AIC_{SM})$, for each combination of ET-demand approach and SM layer. Positive values indicate that including SM reduced AIC and improved model fit. Soil moisture layers were grouped into shallow, root-zone, and deep layers to distinguish near-surface wetness, plant-available water in the rooting zone, and deeper subsurface water storage, respectively.

conditions. The response surface consequently showed the most positive predicted ET errors under the combination of high VPD and low root-zone soil moisture, while predicted errors were less positive or more negative under wetter root-zone conditions at comparable VPD (Fig. 8B).

Sensitivity analyses addressed reuse of the same soil-moisture measurement for multiple ET dates (Supplementary

Fig. S2; Supplementary Table S8). Aggregating records by soil-moisture measurement date reduced the dataset from 297 method-level records to 177 soil-moisture-date-by-method units, while retaining one ET date for each soil-moisture date yielded 59 independent date-level pairs, represented by 177 method-level records. Under both restrictions, several simple correlations moved towards zero. Nevertheless, root-zone

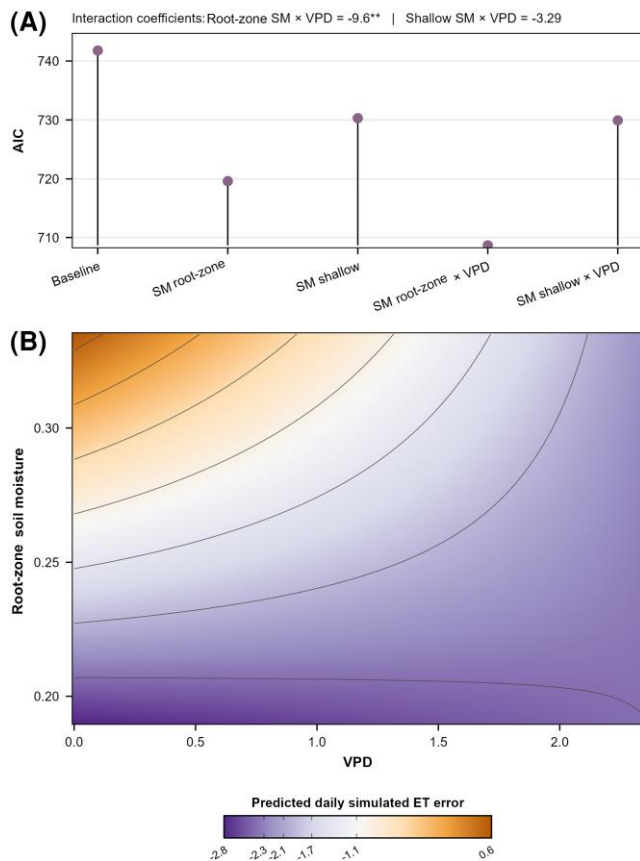


Figure 8 Interactive effects of VPD and soil moisture on ET simulation error in the water-limited no-rain subset. (A) Model comparison based on AIC for the baseline model, models including root-zone or shallow SM as main effects, and models including interaction terms between VPD and SM. The model with the VPD × root-zone SM interaction had the lowest AIC, indicating the strongest support among the tested formulations. (B) Predicted response surface of ET simulation error as a function of VPD and root-zone soil moisture, derived from the interaction model. Higher values indicate more positive predicted ET errors, whereas lower values indicate more negative errors.

soil moisture remained positively associated with ET_{diff} and the VPD × root-zone soil-moisture interaction remained negative. The soil-moisture results were therefore directionally consistent but only partially robust in magnitude.

4. Discussion

4.1 Regime-dependent meteorological attribution

The meteorological attribution of ET simulation errors differed between the two water-supply subsets. Under fully irrigated conditions, ET error was more clearly organized by atmospheric demand and aerodynamic forcing, with conditional contributions from wind speed, VPD, radiation, and the component of maximum temperature not linearly explained by VPD (Nouri et al. 2022). This result indicates that with adequate soil water, model-observation mismatches

become more directly associated with meteorological forcing (Kimball et al. 2023).

Our finding also helps explain why method separation became more evident under irrigation. Penman-Monteith-based methods explicitly represent aerodynamic exchange and humidity effects (Pereira et al. 2015), whereas Priestley-Taylor relies on a fixed empirical coefficient and lacks direct sensitivity to wind speed and VPD (Su and Singh 2023). Such structural differences are likely to be amplified on hot, dry, and windy days, when evaporative demand is large and ET becomes increasingly controlled by turbulent exchange and advection (Chen et al. 2020, Su and Singh 2023). Under full or near-full water supply, these atmospheric controls can be expressed without strong soil water limitation, potentially exposing differences among formulations in their representation of aerodynamic forcing. As crop growth progresses, increasing roughness and a larger transpiration fraction may further strengthen these controls, making ET estimates more sensitive to how formulations represent atmosphere-surface coupling (Ding et al. 2025). By contrast, under water-limited conditions, the conditional explanatory role of meteorological drivers was weaker and more selective, indicating that atmospheric forcing alone was insufficient to account for ET errors. This pattern is consistent with stronger buffering by soil and plant water limitations (Kimball et al. 2023), but it may also reflect limitations in how current ET formulations and crop models represent the soil-plant-atmosphere continuum (Zhao et al. 2022). In many models, the linkage between VPD and soil water stress is represented through empirical and often separable multiplicative stress functions, rather than through an explicit plant hydraulic pathway. As a result, models may generate compensating errors, for example, by underestimating ET sensitivity to atmospheric dryness while over-amplifying sensitivity to soil water deficit (Liu et al. 2020b). However, ET-demand classes were not isolated structural treatments: the models also differed in canopy development, soil-water balance, stress parameterization, and calibration. The observed contrasts therefore cannot be attributed exclusively to the ET-demand equation, and equifinality or compensating calibration effects may also contribute (Wallach et al. 2024). The stage-alignment sensitivity analysis broadly retained the method ordering during early vegetative growth and full canopy cover, but support for method separation during full canopy cover was weaker and the ordering during senescence was less stable. The stage-specific results should therefore be interpreted as broad developmental patterns rather than precise or universal rankings.

This also implies that the apparent role of temperature cannot be interpreted straightforwardly, because heat and atmospheric dryness co-vary strongly (Lesk et al. 2022); we therefore next separated heat from VPD to test whether temperature exerts an independent influence on ET error.

The dominance of the lack-of-correlation component provides an additional diagnosis of model performance. Lack of correlation accounted for approximately two-thirds to three-quarters of total MSE, indicating that the principal limitation was not only persistent mean bias but also an inability to reproduce the timing and relative day-to-day changes of observed ET. This temporal mismatch could arise from

differences in simulated canopy development, soil–water depletion and rewetting, or the timing of responses to radiation, wind, temperature, and atmospheric dryness; observation uncertainty may also contribute. The decomposition itself cannot distinguish among these mechanisms, and the climatic, stage, and soil–moisture analyses should therefore be interpreted as identifying conditions associated with the temporal mismatch rather than uniquely attributing its cause (Sándor et al. 2023).

4.2 Separating residual maximum-temperature variability from atmospheric dryness

As temperature and VPD are tightly coupled, the apparent role of high temperature in ET error cannot be interpreted directly from raw temperature metrics alone (Kingston et al. 2009, Grossiord et al. 2020). By isolating the component of maximum temperature not linearly explained by VPD, the results showed that residual maximum-temperature variability provided additional conditional information in both water-supply subsets. Under irrigated conditions, residual maximum-temperature variability retained clear explanatory ability, indicating that ET error responded not only to vapour demand but also to thermal forcing not captured by VPD alone. By contrast, under water-limited conditions, the component of maximum temperature not linearly explained by VPD became weaker and less stable, suggesting that water limitation constrained the extent to which additional heat could be translated into evaporative response (Villegas et al. 2015, Ding et al. 2025). It is noted that $T_{\text{max,resid}}$ should not be interpreted as a direct measure of physiological heat stress: it removes only the linear component shared with VPD and may retain variation associated with radiation, boundary-layer development, nonlinear temperature–humidity relationships, or other unmeasured conditions.

However, the residual maximum–temperature association should not be interpreted categorically as evidence of an isolated heat effect, or as evidence that heat is unimportant in water-limited systems (Liu et al. 2020b). Under severe atmospheric demand, actual ET can still increase during dry and hot episodes, particularly where vegetation can access deeper water (Mankin et al. 2019) or where flash–drought conditions (Zhao et al. 2022) intensify evaporative demand faster than surface water constraints suppress it. The observed attribution pattern may therefore reflect both genuine hydrological constraints and limitations in the way models represent the interaction between atmospheric demand and declining soil water. In other words, some models may dampen heat-driven ET responses too early (Liu et al. 2020b) by exaggerating the sensitivity of evaporation and transpiration to declining soil water.

This interpretation has broader methodological implications. Studies that attribute ET variability or ET model error to temperature without separating it from VPD may conflate distinct but strongly coupled signals (Kingston et al. 2009), while models that treat atmospheric and soil water deficits as largely independent may misrepresent how heat is actually transmitted into surface fluxes (Novick et al. 2016). A more realistic interpretation is that independent heat effects become detectable mainly where water availability allows atmospheric forcing to be expressed (Yuan et al. 2025), but may also re-emerge in water-limited systems when deeper water access (Ju

et al. 2026) sustains transpiration during hot dry spells (Webber et al. 2016). This motivates examining soil moisture, and particularly its vertical distribution, as an informative state variable (Garrigues et al. 2015) for explaining why similar atmospheric conditions can be associated with different ET errors. Residuals from the daily mixed-effects models showed positive short-lag temporal autocorrelation, particularly under water-limited conditions, but aggregation into non-overlapping three-observation bins broadly retained the direction of the fixed-effect estimates. The attribution results should therefore be interpreted primarily as directional and conditional associations rather than as precise daily-scale effects.

4.3 Soil moisture as an informative state variable in water-limited ET error

The weaker and less consistent meteorological attribution under water-limited conditions indicates that precipitation and atmospheric variables alone do not adequately characterize the water controls on ET error (Jung et al. 2010, Brutsaert 2017). Precipitation reflects forcing, but not the actual soil water available to sustain evaporation and transpiration (Jia et al. 2021). By explicitly introducing observed soil moisture, the results showed that ET errors in water-limited systems are more closely associated with hydrological state beyond rainfall alone, and that this association depended on soil depth. In the original analysis, root-zone moisture provided a stronger and more consistent explanation of ET error than deep soil moisture and generally more information than shallow-layer moisture, indicating that plant-accessible water was particularly relevant to model performance (Maxwell and Condon 2016, Condon et al. 2020, Kimball et al. 2023). However, stricter matching sensitivities attenuated several simple correlations, so root-zone moisture should be interpreted as an informative hydrological state variable rather than the sole control on ET error.

This depth dependence is physically plausible and diagnostically important. Shallow soil moisture responds rapidly to individual rainfall events and can capture short-lived pulses of surface evaporation, but its predictive value is also limited by how models represent near-surface hydraulic processes (Garrigues et al. 2015). In dryland and semi-arid systems, errors in simulating shallow evaporation may arise not only because these pulses are transient but also because soil structure and near-surface hydraulic conductivity are poorly represented (Ritchie et al. 2009, Fatichi et al. 2020), leading to biased partitioning between evaporation losses and downward water redistribution. Previous study (Saadi et al. 2022) showed that improving the representation of water transport through shrinkage cracks can improve simulated soil moisture in clay soils, illustrating how near-surface processes can affect the hydrological state available to ET calculations. Root-zone soil moisture, by contrast, integrates longer-term water availability and is more directly linked to stomatal regulation (Bonan et al. 2014, Zhao et al. 2022), plant water status, and transpiration capacity. It therefore provides a plausible explanation for why similar atmospheric conditions can be associated with different ET errors under water-limited management.

More fundamentally, soil water should not be viewed as just another factor added to meteorological attribution. It is a state variable that mediates whether and how atmospheric demand can be realized at the surface, and this mediation is physically transmitted through the soil–plant–atmosphere continuum (Jung et al. 2010, Liu et al. 2020a, Humphrey et al. 2021). Observed soil–water status should also be distinguished from simulated soil–moisture error: correlations between ET error and coincident soil–moisture simulation error were generally weak, whereas observed root-zone moisture added information to the ET-error model. Where crop models omit explicit plant and soil hydraulics and instead rely on empirical stress scalars, they risk producing compensating errors that overstate soil water limitation while suppressing ET sensitivity to VPD and heat (Liu et al. 2020b). This may be especially problematic during hot, dry periods, when access to deeper soil water or groundwater can sustain transpiration and allow ET to increase despite surface drying. Although stricter treatment of repeated soil–moisture dates attenuated several bivariate correlations, the positive root-zone main effect and negative VPD $\times$ root-zone soil–moisture interaction were retained. Taken together, the results are consistent with water status filtering the association between atmospheric forcing and ET error, while showing that the direction of the root-zone supply-demand interaction was more stable than the magnitude of the simple correlations. Accurate representation of root-zone and deeper water status is therefore likely to be informative for diagnosing ET model error, but the present analysis does not establish a single underlying physiological mechanism.

4.4 Limitations

We also acknowledge several limitations of this study. First, the analysis was based on only two experimental sites, which limits the extent to which the identified patterns can be generalized across global wheat cropping systems. Site, water-supply regime, wheat type, and ET measurement method were not fully crossed: Avignon contributed water-limited wheat observations, whereas Bushland contributed winter-wheat observations under both water-limited and fully irrigated conditions. The pooled regime comparison should therefore be interpreted as a descriptive contrast between the available datasets rather than as an experimentally isolated effect of irrigation. Second, ET observations differed in continuity, measurement approach and likely measurement uncertainty between sites. In particular, ET measurements at the Texas site were derived from weighing lysimeters and were therefore expected to have relatively small observational error, whereas ET observations at Avignon were themselves subject to greater uncertainty associated with the measurement method. As a result, part of the apparent model–data mismatch, especially at Avignon, may reflect not only model error but also uncertainty in the observational reference. In addition, the soil–moisture matching procedure necessarily reduced sample size and relied on short-window assumptions in no-rain periods. Third, the analysis focused on broad ET-method classes rather than on full model structural disentanglement, so some error patterns may also reflect differences in canopy, soil and stress

parameterizations beyond ET formulation itself. Moreover, the soil–moisture matching procedure reduced the effective sample size and relied on a common 5-day matching window during intervals without recorded precipitation or irrigation. The SM-date aggregation and one-ET-per-SM-date sensitivities retained the principal root-zone and interaction directions but weakened several correlations. Finally, the available observations did not allow a full diagnosis of deeper root-water uptake, plant hydraulic regulation, aerodynamic resistance, canopy temperature dynamics, or the partitioning of ET into soil evaporation and transpiration. Future work should therefore expand this framework to broader multi-site datasets, use harmonized ET and soil–moisture observations, and test ET formulations and soil–plant hydraulic processes more directly. Even with these limitations, our study provides clear evidence that ET simulation error cannot be interpreted independently of hydrological regime, and that root-zone water was an informative state variable for understanding ET error in wheat system simulation.

5. Conclusion

Using a multi-model ensemble of wheat crop models grouped by ET-demand approach, changes in daily simulated ET error were investigated across water-limited and irrigated conditions. Absolute ET errors were larger under irrigated conditions, but the contrast was substantially attenuated after normalization by observed ET; larger absolute errors under irrigation arose primarily from an increased lack-of-correlation component rather than stronger systematic bias, pointing to greater day-to-day mismatch between simulations and observations. Differences among ET-demand approaches and climatic factors tended to be more clearly expressed under irrigation, whereas under water-limited conditions, ET error was less well explained by meteorological forcing alone and more strongly linked to soil water status. Nevertheless, because site, wheat type, measurement system, and water-supply regime were not fully crossed, these findings should not be extrapolated as evidence that all models perform uniformly worse under irrigation. Additional site-level evaluations across a broader range of environments and management conditions are needed to assess the generality of these patterns. Our finding implies that improving ET simulation in crop models will require not only better aerodynamic and energy-balance formulations for irrigated systems, but also a more realistic representation of root-zone water availability and its coupling with plant responses in water-limited systems. Recent progress in land-surface modelling (Li et al. 2021, Yao et al. 2022, Umair et al. 2026) further suggests that incorporating plant hydraulic traits and hydraulics-based stomatal regulation can improve the representation of water-stress effects on ET, offering a promising pathway for future refinement of ET processes in crop models.

Supplementary material

Supplementary material is available at *in silico Plants* online.

Conflicts of interest

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

Funding

C.W. acknowledge funding by Marie Skłodowska Curie Actions (No. 101154967). H.W. and D.C. were funded by the Leibniz Female Professorship Award, Germany (Application number: P102/2020) and by the German Research Foundation, Germany (DFG Project Number: No. 470400637). S.E. and K.S. were USDA Agricultural Research Service Ogallala Aquifer Program, a consortium between the USDA–ARS, Kansas State University, Texas A&M University, Texas Tech University, and West Texas A&M University, under USDA ARS Project (Grant No. 3090–13000-016-00D) number. G.S. from the ECMWF was founded by the European Union funded CONCERTO project (Grant No. 101185000). Y.C. and F.T. were supported by the National Key Research and Development Program of China (No. 2025YFF0812104) and International Partnership Program of Chinese Academy of Sciences (Grant No. 177GJHZ2022052MI). S.A. was founded by the German Research Foundation for the ‘FAIRagro’ project (No. 501899475), the project entitled ‘Supporting Agroecological Transitions of Agroforestry Landscapes using an Innovative Transdisciplinary Assessment Framework’ (AGRECO4CAST). S.A. also acknowledge the PARAM-VC project (funding ID for Germany: 01DQ25001A), under the INDO GERMAN SCIENCE AND TECHNOLOGY CENTRE 2 + 2 call, funded jointly by the Department of Science and Technology (DST) under the Government of India and the Federal Ministry of Research, Technology and Space of Germany (BMFTR), LL-SYSTAIN (Grant Labs-2024-IÖR), and ‘SUSTAINCrop’ (HORIZON-EUROPE, funding ID: 0672EMSM).

Data availability

All data used in this study are published and available at: <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/HDKKAL>.

References

- Allen RG, Pereira LS, Raes D *et al.* *FAO Irrigation and Drainage Paper No. 56*, Vol. 56. Rome: Food and Agriculture Organization of the United Nations, 1998, e156.
- Bonan G, Williams M, Fisher R *et al.* Modeling stomatal conductance in the earth system: linking leaf water-use efficiency and water transport along the soil–plant–atmosphere continuum. *Geosci Model Dev* 2014;**7**:2193–222. <https://doi.org/10.5194/gmd-7-2193-2014>
- Bruckler L, Lafolie F, Doussan C *et al.* Modeling soil-root water transport with non-uniform water supply and heterogeneous root distribution. *Plant Soil* 2004;**260**: 205–24. <https://doi.org/10.1023/B:PLSO.0000030187.33135.b8>
- Brutsaert W. Global land surface evaporation trend during the past half century: corroboration by Clausius–Clapeyron scaling. *Adv Water Resour* 2017;**106**:3–5. <https://doi.org/10.1016/j.advwatres.2016.08.014>
- Chen C, Li D, Li Y *et al.* Biophysical impacts of Earth greening largely controlled by aerodynamic resistance. *Sci Adv* 2020;**6**:eabb1981. <https://doi.org/10.1126/sciadv.abb1981>
- Chen X, Huang Y, Liu M *et al.* Improving crop evapotranspiration estimation in the Yellow River Basin: a modified SEBAL model approach and spatiotemporal trend analysis. *Agric Water Manag* 2025;**319**:109771. <https://doi.org/10.1016/j.agwat.2025.109771>
- Condon LE, Atchley AL, Maxwell RM. Evapotranspiration depletes groundwater under warming over the contiguous United States. *Nat Commun* 2020;**11**:873. <https://doi.org/10.1038/s41467-020-14688-0>
- Cui Y, Jia L, Fan W. Estimation of actual evapotranspiration and its components in an irrigated area by integrating the Shuttleworth–Wallace and surface temperature-vegetation index schemes using the particle swarm optimization algorithm. *Agric For Meteorol* 2021;**307**:108488. <https://doi.org/10.1016/j.agrformet.2021.108488>
- Ding J, Yin J, Engel BA *et al.* Estimating cropland evapotranspiration based on remote sensing models: a global meta-analysis. *Eur J Agron* 2025;**170**:127758. <https://doi.org/10.1016/j.eja.2025.127758>
- Evelt SR, Howell TA, Schneider AD *et al.* Energy and water balance modeling of winter wheat. In: *ASAE Paper No. 94-2022, Presented at the 1994 ASAE International Summer Meeting, Kansas City, Missouri, USA*. St. Joseph, MI: American Society of Agricultural Engineers, 1994.
- Evelt SR, Howell TA, Schneider AD *et al.* The bushland weighing lysimeters: a quarter century of crop ET investigations to advance sustainable irrigation. *Trans ASABE* 2016;**59**: 163–79. <https://doi.org/10.13031/trans.59.11159>
- Fatichi S, Or D, Walko R *et al.* Soil structure is an important omission in Earth System Models. *Nat Commun* 2020;**11**: 522. <https://doi.org/10.1038/s41467-020-14411-z>
- Fisher JB, Melton F, Middleton E *et al.* The future of evapotranspiration: global requirements for ecosystem functioning, carbon and climate feedbacks, agricultural management, and water resources. *Water Resour Res* 2017;**53**:2618–26. <https://doi.org/10.1002/2016WR020175>
- Garrigues S, Olioso A, Calvet J-C *et al.* Evaluation of land surface model simulations of evapotranspiration over a 12-year crop succession: impact of soil hydraulic and vegetation properties. *Hydrol Earth Syst Sci* 2015;**19**:3109–31. <https://doi.org/10.5194/hess-19-3109-2015>
- Gauch HG, Hwang JG, Fick GW. Model evaluation by comparison of model-based predictions and measured values. *Agron J* 2003;**95**:1442–6. <https://doi.org/10.2134/agronj2003.1442>
- Geerts S, Raes D. Deficit irrigation as an on-farm strategy to maximize crop water productivity in dry areas. *Agric Water Manag* 2009;**96**:1275–84. <https://doi.org/10.1016/j.agwat.2009.04.009>

- Grossiord C, Buckley TN, Cernusak LA *et al.* Plant responses to rising vapor pressure deficit. *New Phytol* 2020;**226**: 1550–66. <https://doi.org/10.1111/nph.16485>
- Gupta S, Bandyopadhyay K, Sarangi A *et al.* Reference evapotranspiration estimation using AI: performance evaluation across different climates—a review. *Theor Appl Climatol* 2026;**157**:77. <https://doi.org/10.1007/s00704-025-05934-w>
- Hollander M, Wolfe DA, Chicken E. *Nonparametric Statistical Methods*. Hoboken, NJ: John Wiley & Sons, 2013.
- Howell TA, Steiner J, Schneider A *et al.* Evapotranspiration of irrigated winter wheat—Southern High Plains. *Trans ASAE* 1995;**38**:745–59. <https://doi.org/10.13031/2013.27888>
- Humphrey V, Berg A, Ciais P *et al.* Soil moisture–atmosphere feedback dominates land carbon uptake variability. *Nature* 2021;**592**:65–9. <https://doi.org/10.1038/s41586-021-03325-5>
- Jha SK, Ramatshaba TS, Wang G *et al.* Response of growth, yield and water use efficiency of winter wheat to different irrigation methods and scheduling in North China Plain. *Agric Water Manag* 2019;**217**:292–302. <https://doi.org/10.1016/j.agwat.2019.03.011>
- Jia K, Yang Y, Dong G *et al.* Variation and determining factor of winter wheat water requirements under climate change. *Agric Water Manag* 2021;**254**:106967. <https://doi.org/10.1016/j.agwat.2021.106967>
- Ju Q, Ding Y, Shen T *et al.* Impact of groundwater depth on crop coefficient: an improved evapotranspiration model. *Agric Water Manag* 2026;**323**:110060. <https://doi.org/10.1016/j.agwat.2025.110060>
- Jung M, Reichstein M, Ciais P *et al.* Recent decline in the global land evapotranspiration trend due to limited moisture supply. *Nature* 2010;**467**:951–4. <https://doi.org/10.1038/nature09396>
- Kang S, Payne WA, Evett SR *et al.* Simulation of winter wheat evapotranspiration in Texas and Henan using three models of differing complexity. *Agric Water Manag* 2009;**96**: 167–78. <https://doi.org/10.1016/j.agwat.2008.07.006>
- Katul GG, Oren R, Manzoni S *et al.* Evapotranspiration: a process driving mass transport and energy exchange in the soil–plant–atmosphere–climate system. *Rev Geophys* 2012;**50**: RG3002. <https://doi.org/10.1029/2011RG000366>
- Kimball BA, Thorp KR, Boote KJ *et al.* Simulation of evapotranspiration and yield of maize: an inter-comparison among 41 maize models. *Agric For Meteorol* 2023;**333**:109396. <https://doi.org/10.1016/j.agrformet.2023.109396>
- Kingston DG, Todd MC, Taylor RG *et al.* Uncertainty in the estimation of potential evapotranspiration under climate change. *Geophys Res Lett* 2009;**36**:L20403. <https://doi.org/10.1029/2009GL040267>
- Lesk C, Anderson W, Rigden A *et al.* Compound heat and moisture extreme impacts on global crop yields under climate change. *Nat Rev Earth Environ* 2022;**3**:872–89. <https://doi.org/10.1038/s43017-022-00368-8>
- Li L, Yang ZL, Matheny AM *et al.* Representation of plant hydraulics in the Noah-MP land surface model: model development and multiscale evaluation. *J Adv Model Earth Syst* 2021;**13**:e2020MS002214. <https://doi.org/10.1029/2020MS002214>
- Lian X, Piao S, Chen A *et al.* Multifaceted characteristics of dry-land aridity changes in a warming world. *Nat Rev Earth Environ* 2021;**2**:232–50. <https://doi.org/10.1038/s43017-021-00144-0>
- Lian X, Piao S, Huntingford C *et al.* Partitioning global land evapotranspiration using CMIP5 models constrained by observations. *Nat Clim Chang* 2018;**8**:640–6. <https://doi.org/10.1038/s41558-018-0207-9>
- Liu L, Gudmundsson L, Hauser M *et al.* Soil moisture dominates dryness stress on ecosystem production globally. *Nat Commun* 2020a;**11**:4892. <https://doi.org/10.1038/s41467-020-18631-1>
- Liu Y, Kumar M, Katul GG *et al.* Plant hydraulics accentuates the effect of atmospheric moisture stress on transpiration. *Nat Clim Chang* 2020b;**10**:691–5. <https://doi.org/10.1038/s41558-020-0781-5>
- Lu Y, Yang T, Fu J *et al.* Utility of the standardized precipitation evapotranspiration index (SPEI) to detect agricultural droughts over China. *J Hydrol Reg Stud* 2025;**58**:102190. <https://doi.org/10.1016/j.ejrh.2025.102190>
- Mankin JS, Seager R, Smerdon JE *et al.* Mid-latitude freshwater availability reduced by projected vegetation responses to climate change. *Nat Geosci* 2019;**12**:983–8. <https://doi.org/10.1038/s41561-019-0480-x>
- Maxwell RM, Condon LE. Connections between groundwater flow and transpiration partitioning. *Science* 2016;**353**: 377–80. <https://doi.org/10.1126/science.aaf7891>
- McMillan H. Linking hydrologic signatures to hydrologic processes: a review. *Hydrol Process* 2020;**34**:1393–409. <https://doi.org/10.1002/hyp.13632>
- Milly PC, Dunne KA. Potential evapotranspiration and continental drying. *Nat Clim Chang* 2016;**6**:946–9. <https://doi.org/10.1038/nclimate3046>
- Nouri M, Hoogenboom G, Bannayan M *et al.* CSM-CERES-wheat sensitivity to evapotranspiration modeling frameworks under a range of wind speeds. *Water (Basel)* 2022;**14**:3023. <https://doi.org/10.3390/w14193023>
- Novick KA, Ficklin DL, Stoy PC *et al.* The increasing importance of atmospheric demand for ecosystem water and carbon fluxes. *Nat Clim Chang* 2016;**6**:1023–7. <https://doi.org/10.1038/nclimate3114>
- Oki T, Kanae S. Global hydrological cycles and world water resources. *Science* 2006;**313**:1068–72. <https://doi.org/10.1126/science.1128845>
- Pereira LS, Allen RG, Smith M *et al.* Crop evapotranspiration estimation with FAO56: past and future. *Agric Water Manag* 2015;**147**:4–20. <https://doi.org/10.1016/j.agwat.2014.07.031>
- Piao S, Wang X, Park T *et al.* Characteristics, drivers and feedbacks of global greening. *Nat Rev Earth Environ* 2020;**1**: 14–27. <https://doi.org/10.1038/s43017-019-0001-x>
- Qiu R, Katul GG, Zhang L *et al.* The effects of changing environments, abiotic stresses, and management practices on cropland evapotranspiration: a review. *Rev Geophys*

- 2025;63:e2024RG000858. <https://doi.org/10.1029/2024RG000858>
- Ritchie JT, Porter CH, Judge J *et al.* Extension of an existing model for soil water evaporation and redistribution under high water content conditions. *Soil Sci Soc Am J* 2009;73:792–801. <https://doi.org/10.2136/sssaj2007.0325>
- Rosenzweig C, Jones JW, Hatfield JL *et al.* The agricultural model intercomparison and improvement project (AgMIP): protocols and pilot studies. *Agric For Meteorol* 2013;170:166–82. <https://doi.org/10.1016/j.agrformet.2012.09.011>
- Saadi S, Pattey E, Jégo G *et al.* Prediction of rainfed corn evapotranspiration and soil moisture using the STICS crop model in eastern Canada. *Field Crops Res* 2022;287:108664. <https://doi.org/10.1016/j.fcr.2022.108664>
- Sándor R, Ehrhardt F, Grace P *et al.* Residual correlation and ensemble modelling to improve crop and grassland models. *Environ Model Softw* 2023;161:105625. <https://doi.org/10.1016/j.envsoft.2023.105625>
- Seneviratne SI, Corti T, Davin EL *et al.* Investigating soil moisture–climate interactions in a changing climate: a review. *Earth Sci Rev* 2010;99:125–61. <https://doi.org/10.1016/j.earscirev.2010.02.004>
- Su Q, Singh VP. Calibration-free Priestley-Taylor method for reference evapotranspiration estimation. *Water Resour Res* 2023;59:e2022WR033198. <https://doi.org/10.1029/2022WR033198>
- Trnka M, Rötter RP, Ruiz-Ramos M *et al.* Adverse weather conditions for European wheat production will become more frequent with climate change. *Nat Clim Chang* 2014;4:637–43. <https://doi.org/10.1038/nclimate2242>
- Umair M, Melton JR, Roy A *et al.* Implementing a plant hydraulics parameterization in the Canadian land surface scheme including biogeochemical cycles (CLASSIC) v. 1.4. *J Adv Model Earth Syst* 2026;18:e2024MS004385. <https://doi.org/10.1029/2024MS004385>
- Vereecken H, Amelung W, Bauke SL *et al.* Soil hydrology in the earth system. *Nat Rev Earth Environ* 2022;3:573–87. <https://doi.org/10.1038/s43017-022-00324-6>
- Villegas JC, Dominguez F, Barron-Gafford GA *et al.* Sensitivity of regional evapotranspiration partitioning to variation in woody plant cover: insights from experimental dryland tree mosaics. *Glob Ecol Biogeogr* 2015;24:1040–8. <https://doi.org/10.1111/geb.12349>
- Wallach D, Buis S, Seserman D-M *et al.* A calibration protocol for soil-crop models. *Environ Model Softw* 2024;180:106147. <https://doi.org/10.1016/j.envsoft.2024.106147>
- Wang X, Müller C, Elliot J *et al.* Global irrigation contribution to wheat and maize yield. *Nat Commun* 2021;12:1235. <https://doi.org/10.1038/s41467-021-21498-5>
- Wanniarachchi S, Sarukkalgige R. A review on evapotranspiration estimation in agricultural water management: past, present, and future. *Hydrology* 2022;9:123. <https://doi.org/10.3390/hydrology9070123>
- Webber H, Cooke D, Wang C *et al.* Wheat crop models underestimate drought stress in semi-arid and Mediterranean environments. *Field Crops Res* 2025;332:110032. <https://doi.org/10.1016/j.fcr.2025.110032>
- Webber H, Gaiser T, Oomen R *et al.* Uncertainty in future irrigation water demand and risk of crop failure for maize in Europe. *Environ Res Lett* 2016;11:074007. <https://doi.org/10.1088/1748-9326/11/7/074007>
- Wu X, Xu Y, Shi J *et al.* Estimating stomatal conductance and evapotranspiration of winter wheat using a soil-plant water relations-based stress index. *Agric For Meteorol* 2021;303:108393. <https://doi.org/10.1016/j.agrformet.2021.108393>
- Yang Y, Anderson MC, Gao F *et al.* Phenological corrections to a field-scale, ET-based crop stress indicator: an application to yield forecasting across the US Corn Belt. *Remote Sens Environ* 2021;257:112337. <https://doi.org/10.1016/j.rse.2021.112337>
- Yang Y, Roderick ML, Guo H *et al.* Evapotranspiration on a greening Earth. *Nat Rev Earth Environ* 2023;4:626–41. <https://doi.org/10.1038/s43017-023-00464-3>
- Yao Y, Joetzier E, Ciais P *et al.* Forest fluxes and mortality response to drought: model description (ORCHIDEE-CAN-NHA, r7236) and evaluation at the Caxiuanã drought experiment. *Geosci Model Dev Discuss* 2022;15:7809–33. <https://doi.org/10.5194/gmd-15-7809-2022>
- Yuan W, Tian J, Wang M *et al.* Impacts of rising atmospheric dryness on terrestrial ecosystem carbon cycle. *Nat Rev Earth Environ* 2025;6:712–27. <https://doi.org/10.1038/s43017-025-00726-2>
- Zaerpour M, Hatami S, Ballarin AS *et al.* Agriculture's impact on water–energy balance varies across climates. *Proc Natl Acad Sci U S A* 2025;122:e2410521122. <https://doi.org/10.1073/pnas.2410521122>
- Zhao J, Sun S, Yin Y *et al.* Advancing evapotranspiration modeling with optimized soil and canopy resistance combinations. *Water Resour Res* 2025;61:e2024WR039252. <https://doi.org/10.1029/2024WR039252>
- Zhao M, Geruo A, Liu Y *et al.* Evapotranspiration frequently increases during droughts. *Nat Clim Chang* 2022;12:1024–30. <https://doi.org/10.1038/s41558-022-01505-3>