UC Merced

UC Merced Previously Published Works

Title

Multi-scale assessment of wavelet-guided deep learning for reference evapotranspiration forecasting to support vineyard water management

Permalink

https://escholarship.org/uc/item/4v36n611

Journal

Agricultural Water Management, 338

ISSN

0378-3774

Authors

Gao, Rui
Safeeq, Mohammad
Abatzoglou, John T
et al.

Publication Date

2026-12-01

DOI

10.1016/j.agwat.2026.110892

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at https://creativecommons.org/licenses/by/4.0/

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Agricultural Water Management 338 (2026) 110892

Contents lists available at ScienceDirect

Agricultural Water Management

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

Multi-scale assessment of wavelet-guided deep learning for reference evapotranspiration forecasting to support vineyard water management

Rui Gao [a,*], Mohammad Safeeq [a], John T. Abatzoglou [b], Joshua H. Viers [a]

[a] Civil and Environmental Engineering Department, University of California, Merced, CA, 95343, USA
[b] Management of Complex Systems Department, University of California, Merced, CA, 95343, USA

ARTICLE INFO

Keywords:
Continuous wavelet transform
Time-frequency decomposition
Input sequence length
CNN-LSTM
Energy-balance closure
Actual crop coefficient

ABSTRACT

Reliable multi-week forecasting of grass reference evapotranspiration (ET_o) can provide advance information on atmospheric water demand for agricultural water management, yet its skill is often judged using conventional accuracy metrics that may be inflated by the strong seasonality of ET_o. Here, we develop a convolutional neural network-long short-term memory (CNN-LSTM) framework for daily ET_o forecasting at lead times of up to eight weeks in Madera County, California, and evaluate it across regional and vineyard scales against climatology and persistence forecasts. We use continuous wavelet transform (CWT) analysis to identify a dominant 93-day timescale in regional ET_o and test whether this timescale can guide the design of input sequence lengths. A 60-day input outperformed longer inputs, including the CWT-derived timescale, indicating that the dominant timescale of ET_o variability does not necessarily correspond to the timescale over which ET_o is predictable. The CNN-LSTM reproduced the seasonal evolution of ET_o and corrected the seasonal bias of its persistence baseline, but it did not significantly outperform climatology at any lead, and its skill in predicting ET_o anomalies was confined to the first two days. Coupling the ET_o forecasts with actual crop coefficient forecasts to estimate vineyard water use introduced further uncertainty, and the evaluated performance depended on the choice of eddy covariance reference in ways that differed between vineyards under contrasting water stress. Our results suggest that benchmark-relative evaluation is essential for strong seasonal variables such as ET_o, and that improving multi-week forecasts of ET_o and crop water use will require additional predictors, better characterization of the uncertainty in gridded ET_o products, and improved crop coefficient estimation.

1. Introduction

Water scarcity is increasing the need for decision-support tools that enable proactive irrigation planning, particularly where operational decisions require lead times of several weeks. This need is especially relevant in California's high-value perennial agriculture, where irrigation planning under a changing climate depends on understanding how phenology and biophysical processes translate into crop water use (Kandhway et al., 2025). More broadly, California is facing intensifying hydroclimatic pressures that increase the urgency of anticipating water

Abbreviations: EC, Eddy covariance; CIMIS, California Irrigation Management Information System; GRAPEX, Grape Remote sensing Atmospheric Profile and Evapotranspiration eXperiment; ET, Evapotranspiration; ET_o, Grass reference evapotranspiration; ET_o^P, Grass reference evapotranspiration extracted from the spatial CIMIS maps (point-based); ET_o^{PC}, Point-based grass reference evapotranspiration extracted from the spatial CIMIS maps and corrected using a site-specific correction factor derived from local comparison; $ET_{c\ act}$, Actual crop evapotranspiration; PM, Penman-Monteith; PM-ET_o, Reference evapotranspiration calculated using the Penman-Monteith equation; H-ET_o, Reference evapotranspiration calculated using the Hargreaves equation; $ET_{c\ act}^O$, OpenET ensemble actual evapotranspiration; $ET_{c\ act}^{EC}$, Eddy covariance flux tower monitored actual evapotranspiration; $ET_{c\ act}^{EC,BR}$, Bowen-ratio corrected actual evapotranspiration based on eddy-covariance flux tower estimates; LRU, Lead-Ratio Updating, an empirical method for variable forecasting; $ET_{c\ act}^{LRU}$, Lead-Ratio Updating method for actual evapotranspiration forecasting; $K_{c\ act}$, Actual crop coefficient; $K_{c\ act}^{LRU}$, Lead-Ratio Updating an empirical method for; $K_{c\ act}$, forecasting; AOI, Area of interest; MOI, Modeling area of interest; CNN-LSTM, Convolutional neural network-long short-term memory; CWT, Continuous wavelet transform; R2, Coefficient of determination; RMSE, Root mean square error; PBIAS, Percent bias; SS, Mean squared error skill score; SS^{clim}, Mean squared error skill score of the CNN-LSTM relative to the climatological forecast; SS^{pers}, Mean squared error skill score of the CNN-LSTM relative to the anomaly persistence forecast; ACC, Anomaly correlation coefficient; L_{in}, Input sequence length.

* Corresponding author.
E-mail address: rui.ray.gao@gmail.com (R. Gao).

https://doi.org/10.1016/j.agwat.2026.110892
Received 4 June 2026; Received in revised form 6 October 2026; Accepted 22 October 2026
Available online 7 November 2026
0378-3774/© 2026 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

demand and ecosystem stress. Sierra Nevada headwater watersheds, which are a vital source of water for ecosystems and the economy, are increasingly threatened by warming-driven drought stress and wildfire risk (Abatzoglou and Williams, 2016; Casirati et al., 2025), while recent work further suggests that warming can intensify vegetation water stress and alter water loss under changing climatic conditions (Casirati et al., 2026). In this context, evapotranspiration (ET), the combined flux of evaporation and transpiration (Jensen and Allen, 2016), naturally emerges as a key forecasting variable not only for irrigation planning and agricultural system management, but also for the conservation of wetlands and forest habitats (Granata and Di Nunno, 2021).

Current ET forecasting efforts generally follow two main pathways: direct forecasting of actual crop ET ($ET_{c\ act}$) and forecasting of grass reference ET (ET_o), which can subsequently be combined with the classical crop-coefficient method (Allen et al., 2005) to estimate $ET_{c\ act}$ by coupling the actual crop coefficient ($K_{c\ act}$) and ET_o (Fernández, 2023; Pereira et al., 2021b). Although direct $ET_{c\ act}$ forecasting is more closely linked to realized crop water use, it is also more difficult because it must account for both atmospheric demand and time-varying crop and land-surface conditions. Capturing these controls often requires more advanced techniques and more demanding inputs that resolve land-surface characteristics at fine spatial and temporal scales, such as land surface temperature (Anderson et al., 2024). In contrast, ET_o is governed primarily by meteorological forcing and provides a standardized measure of atmospheric evaporative demand above a standardized reference surface (Jensen and Allen, 2016; Pereira et al., 2021a), making it less dependent on site-specific crop, soil, and management conditions and therefore more transferable and scalable as a forecasting target than $ET_{c\ act}$. Similar conclusions have been reported in point-based ET_o forecasting using machine learning, including deep learning, at hourly (Ahmadi et al., 2024), daily (Pino-Vargas et al., 2022), and monthly scales (Ahmadi et al., 2023). However, before ET_o forecasts can be used to estimate $ET_{c\ act}$ through the crop coefficient approach, important gaps remain in the design and evaluation of deep learning frameworks for ET_o forecasting.

Recent studies have demonstrated that deep learning can achieve accurate ET_o forecasting over short forecast horizons, such as 7 days (e.g., Ben Hamouda et al., 2022; Gómez et al., 2024). However, these studies largely evaluated model performance under short-range or otherwise fixed forecast settings, without systematically addressing how forecasting skill changes as the forecast horizon extends. From an operational perspective, Shukla et al. (2015) identified around one month as a practical threshold for actionable forecasts, and similar patterns have been reported for specialty-crop climate services in California, where forecast skill is strongest at approximately one month and declines at longer horizons (Jha and Pathak, 2024). In related applied forecasting contexts, sub-seasonal forecasts at 2- to 4-week lead times have also been recognized as useful for proactive planning and resource allocation (Abatzoglou et al., 2023). Together, these studies highlight the need to evaluate ET_o forecasting skill across continuous multi-week lead times, rather than at a single fixed horizon.

Yet an important design question remains unresolved: how much historical information, or the input sequence length (L_{in}), is needed for a deep learning model to support a given forecast horizon (Gao et al., 2025)? Most existing studies report forecasting skill under fixed combinations of L_{in} and forecast horizon, rather than explicitly addressing the broader relationship between the two. In addition, the spatial scale of assessment also matters, because water scarcity patterns and their controlling factors can differ substantially between local and regional analyses, limiting the applicability of management strategies developed from a single scale (Li et al., 2024; Liu et al., 2025). These unresolved issues highlight the need for ET_o forecasting frameworks that jointly consider input-window design, forecast-horizon dependence, and forecast transferability across decision-relevant spatial scales, thereby supporting both field-scale crop water decisions and regional water allocation.

Our previous work showed that spatial ET_o at 2-km pixel resolution from the California Irrigation Management Information System (CIMIS) was sufficient to support skillful deep-learning-based forecasting, whereas incorporating additional spatial CIMIS variables did not yield clear improvement (Gao et al., 2025). CIMIS is a foundational public source of weather and ET_o information for irrigation scheduling and crop water allocation in California (Parker and Zilberman, 1996; Trilnick et al., 2022). These findings suggest that, rather than further expanding model inputs, an important next step is to examine the intrinsic temporal structure of ET_o itself. Because ET_o reflects the integrated response of atmospheric evaporative demand to meteorological forcing, its time series may contain dominant temporal scales that are relevant to forecasting-system design. In this context, continuous wavelet transform (CWT), which can identify localized, scale-dependent temporal variability in non-stationary signals (Altunkaynak and Ozger, 2016; Sharma et al., 2024), provides a useful framework for addressing this design question. More specifically, it raises the question of whether a dominant ET_o timescale can provide a physically informed reference point for selecting L_{in}, even if it does not directly determine the optimal model configuration.

Our objectives were to: (1) identify dominant temporal scales in ET_o and evaluate whether these scales provide useful guidance for selecting the historical L_{in}, (2) quantify how ET_o forecast skill evolves across extended lead times and spatial scales, and (3) examine, as an extension, how empirical $K_{c\ act}$ forecasting influences coupled $ET_{c\ act}$ performance and reveals the main limitations of extending strong ET_o forecasts to crop-coefficient-based $ET_{c\ act}$ forecasting. By explicitly linking timescale, L_{in}, and forecast horizon, this study provides practical guidance for the development of multi-week ET_o forecasting models for agricultural water management, while also showing why reliable extension to coupled $ET_{c\ act}$ estimation remains constrained by uncertainty in $K_{c\ act}$ estimation.

2. Materials and methods

2.1. Study area and data preparation

This study was conducted in Madera County, located within the San Joaquin Valley, California, USA. The southwestern portion of the county is dominated by agriculture, whereas the northeastern region is characterized by mountainous natural landscapes with relatively low human disturbance. Elevation decreases from the northeastern uplands toward the agricultural lowlands in the southwest (Fig. 1). Two vineyard sites instrumented through the GRAPEX program (Kustas et al., 2019, 2022) and located less than 2 km apart were used for field-scale evaluation: RIP720, a Merlot vineyard, and RIP760, a Chardonnay vineyard. RIP720 is an approximately 400 m by 400 m vineyard block that included four management treatments, each monitored by an eddy covariance (EC) tower (Gao et al., 2023; Gao et al., 2026; Safre et al., 2022): two treatments were subjected to severe water stress and two to normal deficit irrigation. The measured energy-balance components from the four towers were averaged to provide a field-scale representation of RIP720. By contrast, RIP760 was maintained under little to no water stress, so that the two sites represented contrasting water-stress conditions. In addition to the county-scale Area of Interest (AOI), we defined a Modeling Area of Interest (MOI) as a rectangular domain that encompasses the primary agricultural region of Madera County (Fig. 1).

Three primary datasets were used in the forecasting and evaluation framework. First, daily gridded ET_o was obtained from the CIMIS following Gao et al. (2026d). The CIMIS ET_o was available at 2-km spatial resolution for 2014–01–01–2018–12–31 and served as the target variable for regional CNN-LSTM forecasting. Second, daily OpenET ensemble $ET_{c\ act}$, denoted as $ET^{O}_{c\ act}$, was obtained at two vineyard sites for 2017–2018 using the method described by Gao et al. (2026c). Third, the EC flux tower observations collected at RIP720

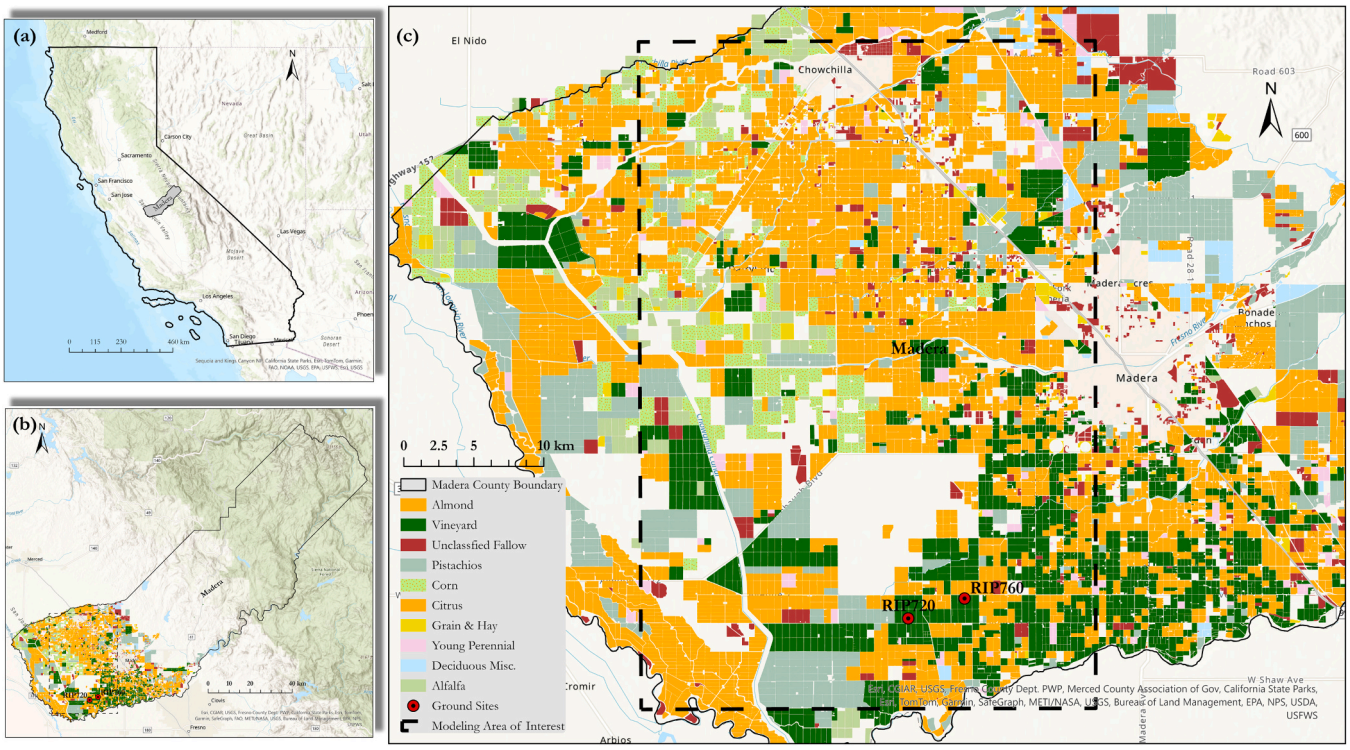

Fig. 1. Study region and modeling domain for ET_o forecasting in Madera County, California. Panel (a) shows the location of Madera County within California. Panel (b) shows the distribution of agricultural lands within the county. Panel (c) shows 2018 Land IQ land-use/land-cover classes across Madera County, with the county boundary overlaid on a hillshade basemap. The black dashed rectangle indicates the modeling area of interest (MOI) used as the spatial input domain for CNN-LSTM ET_o forecasting. The eddy covariance flux tower sites, RIP720 and RIP760, are indicated by red circles with black dots at their centers.

and RIP760 during 2018 were used to derive tower-based $ET_{c\ act}$ for field-scale evaluation (Fig. 2a,c). EC towers were also instrumented with meteorological sensors at 3.5 m above ground level (Nassar et al., 2021). We adjusted wind speed to the standard 2-m height, while relative humidity and air temperature were assumed to be unchanged between the two heights. These meteorological observations were used to calculate local ET_o using a Python implementation (Gao, 2026a) of the Penman-Monteith (PM) equation (Allen et al., 2005), hereafter denoted

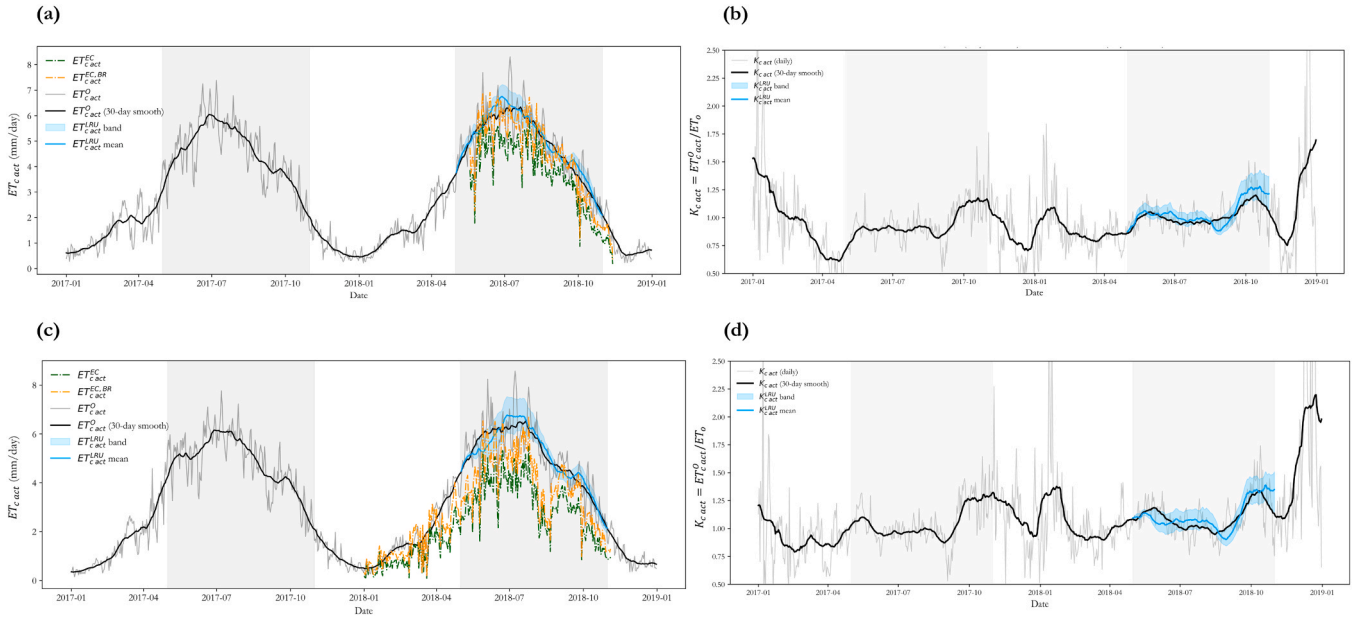

Fig. 2. Panels (a) and (c) show daily OpenET ensemble $ET_{c\ act}$ ($ET^{O}_{c\ act}$, gray) and its 30-day smoothed series (black) at RIP720 and RIP760, respectively. Direct $ET_{c\ act}$ observations from the EC flux tower ($ET^{EC}_{c\ act}$ in dark-green dash-dot line) and the Bowen-ratio corrected $ET_{c\ act}$ labeled as $ET^{EC,BR}_{c\ act}$ in dark-orange dash-dot line. The blue curves and shaded envelopes represent the mean ($ET^{LRU}_{c\ act}$, blue) and range of the empirical 56-day $ET^{LRU}_{c\ act}$ during the 2018 growing season. Panels (b) and (d) show the corresponding $K_{c\ act}$ time series, including the daily and 30-day smoothed $K_{c\ act}$ from OpenET-CIMIS and the empirical forecast $K_{c\ act}$ ($K^{LRU}_{c\ act}$). Light gray shading highlights the May-October periods in 2017 and 2018.

as PM-ET_o, and the Hargreaves equation (Hargreaves and Allen, 2003), hereafter denoted as H-ET_o.

$$EBR = \frac{H + LE}{R_n - G} \tag{1}$$

Energy-balance closure was explicitly considered when processing the EC observations. Closure was assessed using the energy-balance ratio (EBR, Eq. 1). The median EBR was approximately 0.83 at RIP720 and 0.76 at RIP760, indicating that the measured turbulent flux did not fully account for the available energy at either site (Fig. A1 and A3). A Bowen-ratio correction was applied by allocating the residual available energy between sensible heat flux (H) and latent heat flux (LE) according to their measured ratio. This processing produced two EC-based observational references: directly measured, denoted as $ET^{EC}_{c\ act}$, and Bowen-ratio corrected, denoted as $ET^{EC,BR}_{c\ act}$. Both references were retained to evaluate the sensitivity of field-scale forecast performance to energy-balance closure. Fig. A 1 and Fig. A 3 present the EBR distributions, hourly comparisons between direct and corrected LE ($W \bullet m^{-2}$), and the corresponding daily ET ($mm \bullet day^{-1}$) time series for RIP720 and RIP760, respectively.

2.2. Methods

The methodological workflow was designed to evaluate regional ET_o forecasting and its site-scale applicability. First, CWT analysis was applied to the CIMIS ET_o to identify dominant temporal scales and guide the selection of L_{in} for CNN-LSTM forecasting. Second, the CNN-LSTM model was trained to generate daily ET_o forecasts over a 56-day forecast horizon, and model skill was evaluated as a function of forecast lead time and L_{in}. The forecast horizon was set to 56 days to represent eight complete weekly lead intervals, enabling forecast degradation to be assessed consistently across decision-relevant weekly blocks while preserving daily lead-specific diagnostics. Because approximately one month, or four weeks, is often considered an actionable forecasting horizon in agricultural planning, the 56-day design also allowed us to examine whether ET_o forecast skill persisted beyond that benchmark.

A point-based ET_o was extracted from CIMIS ET_o for each vineyard, hereafter denoted as CIMIS-ET^P_o, to represent its practical use in irrigation management. We then compared CIMIS-ET^P_o values with PM-ET_o to assess field-scale representativeness and the potential need for local correction, and the corresponding correction was hereafter denoted as CIMIS-ET^{PC}_o. Finally, the forecasted ET_o was used in an additional downstream $ET_{c\ act}$ estimation step through empirical $K_{c\ act}$ coupling to illustrate the implications of ET_o forecast and correction uncertainty for vineyard water-use assessment.

2.2.1. Spatial wavelet analysis of reference evapotranspiration and maximum-period sensitivity

We applied CWT analysis (Fig. 3) to the CIMIS ET_o time series from 2014–01–01 to 2017–12–31 to identify dominant periods from weekly to seasonal scales. The CWT of a time series $x(t)$ is given in Eq. 2, where ψ is the mother wavelet, a is the wavelet scale parameter, and b is the translation parameter. The scale parameter a controls the temporal width of the wavelet and thus the characteristic period being analyzed, while the translation parameter b shifts the wavelet along the time axis. The resulting coefficients $W(a,b)$ describe localized temporal structure across time and period (Adamowski et al., 2009). Wavelet power was calculated as Eq. 6, where $P(a,b)$ represents the strength of variability at period a and time b.

$$W(a,b) = \frac{1}{\sqrt{a}} \int x(t)\psi^*\left(\frac{t-b}{a}\right)dt \tag{2}$$

$$\psi(t) = \pi^{-1/4} \bullet e^{i\omega_0 t} \bullet e^{-t^2/2} \tag{3}$$

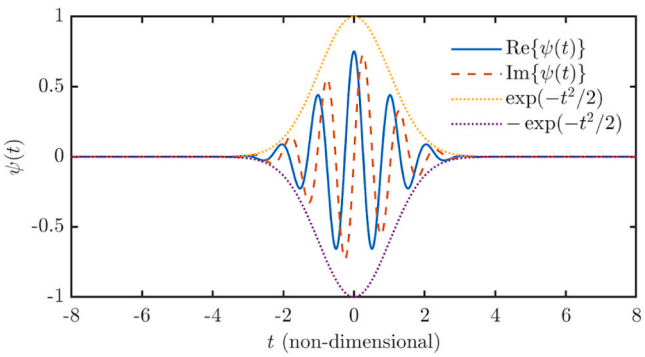

Fig. 3. Illustration of the Morlet mother wavelet $\psi(t)$, used in the CWT. The figure shows the real part (solid blue line) and imaginary part (dashed red line) of the wavelet as functions of non-dimensional time t. The oscillatory carrier is localized by the Gaussian envelope $e^{\frac{-t^2}{2}}$ (dotted orange line) and its reflected counterpart (dotted purple line), which bound the wavelet amplitude. The Morlet wavelet was used in this study to decompose CIMIS ET_o time series and identify localized temporal variability across periods.

$$Re\{\psi(t)\} = \pi^{-1/4} \bullet \cos(\omega_0 t) \bullet e^{-t^2/2} \tag{4}$$

$$Im\{\psi(t)\} = \pi^{-1/4} \bullet \sin(\omega_0 t) \bullet e^{-t^2/2} \tag{5}$$

$$P(a,b) = |W(a,b)|^2 \tag{6}$$

We used the complex Morlet wavelet defined in Eq. 3, with central frequency $\omega_0 = 6$, consistent with the default configuration in the *pycwt* package. Its real and imaginary components form cosine- and sine-modulated Gaussian functions (Eq. 4 and Eq. 5), allowing the wavelet coefficients to retain both amplitude and phase information.

For each 2-km pixel within the AOI, wavelet power was averaged over time to obtain the global wavelet power spectrum. The dominant period was defined as the period with the maximum global wavelet power within the analyzed period band. The minimum analyzed period was fixed at 7 days, whereas the maximum analyzed period, P_{max}, was varied from 14 to 364 days in 7-day increments to evaluate sensitivity to the upper bound of the period band. For each P_{max}, the dominant period was recomputed for every pixel within the restricted interval of 7 days to P_{max}.

Each sensitivity run was evaluated using (1) the number of unique dominant periods across the domain after rounding to integer days, (2) the fraction of pixels within the fixed MOI assigned to the candidate dominant periods, and (3) the spatial correspondence of the resulting dominant-period map with agricultural and natural landscape gradients in Madera County. These diagnostics were used to assess the robustness of candidate dominant periods across P_{max} settings and to select an upper bound that produced a stable and spatially interpretable dominant-period pattern.

The wavelet-derived dominant period informed the selection of candidate historical L_{in} for the CNN-LSTM experiments. Input windows shorter than, approximately equal to, and longer than this wavelet-derived dominant period were evaluated to determine whether the characteristic timescale also provided an effective historical context for multi-lead forecasting.

2.2.2. Reference evapotranspiration forecasting

2.2.2.1. Deep learning model for multi-lead reference evapotranspiration forecasting. We developed the CNN-LSTM model over the MOI (Fig. 1), a 19 × 13 grid at 2-km resolution covering the primary agricultural region, rather than the full 59 × 71 county-wide domain. This spatial extent was selected because (1) both evaluation sites are located within

the MOI and (2) its smaller, rectangular boundary reduces the computational cost of model training and simplifies boundary handling by the convolutional layers, particularly because pixels outside the county boundary were masked as missing values in the full domain. The CIMIS ET_o record from 2014–01–01 to 2018–12–31 was divided chronologically into training (2014–01–01–2017–12–31), validation (2018–05–01–2018–07–31), and testing (2018–08–01–2018–10–31) periods. Because the CNN-LSTM requires a L_{in}, the number of valid samples within a fixed calendar window decreases as L_{in} increases. To ensure comparable evaluation across experiments, the validation and testing calendar windows were kept fixed, while the number of valid sequences was allowed to vary with L_{in}. Forecasts were generated over a 56-day forecast horizon and evaluated at 1-day lead increments.

Each daily CIMIS ET_o was first encoded using a single 2-D convolutional layer (3 × 3 kernel) followed by 2 × 2 max pooling to obtain a compact spatial representation. These encoded features were flattened and passed to a single-layer LSTM to capture temporal dependencies across the input sequence. The final LSTM hidden state was then decoded through a fully connected layer reshaped into a feature map, followed by a 1 × 1 convolution to generate daily ET_o forecasts for lead days 1–56.

We used a residual-learning formulation with a persistence baseline in the CNN-LSTM model to stabilize the multi-lead ET_o forecasting. The last observed ET_o map in the input sequence was repeated across all lead times to define the baseline forecast (hereafter raw persistence), and the model learned lead-specific residual fields relative to that baseline. These residuals were constrained using a scaled hyperbolic tangent function before being added back to the persistence baseline, allowing the model to learn deviations from recent ET_o conditions while limiting unrealistic departures at longer lead times. Alternative loss formulations, including reweighted losses for extreme ET_o values and a weighted Huber loss (Gupta et al., 2020), were also tested, but they did not yield clear improvement over the standard mean squared error objective. The final model was therefore trained with the Adam optimizer and L2 weight decay to minimize the mean squared error loss, with gradient clipping applied to improve numerical stability.

We implemented early stopping on validation loss (patience=5; minimum improvement threshold=1×10^{-4}) and retained the best-performing checkpoint for each configuration. To reduce the risk of under- or over-parameterization, candidate configurations were filtered using an estimated input-to-parameter ratio based on the number of training samples, input sequence length, the number of valid pixels in the MOI, and the total number of model parameters; configurations below the minimum ratio threshold (5 in this study) were excluded from further evaluation. Forecast skill was evaluated across lead days 1–56.

2.2.2.2. Climatology-based forecasting. Climatological forecasts (Murphy, 1992), a lead-time-independent method, were used as the first baseline for ET_o forecasting. For each grid cell, (i,j) and target calendar day (d, denotes the day of year in the 2018 test year), the climatological forecast, $\widehat{ET}^{clim}_{o,i,j,d}$ (Eq. 7), was calculated as the mean ET_o for the corresponding day across the training years (y).

$$\widehat{ET}^{clim}_{o,i,j,d} = \frac{1}{N}\sum_{y=1}^{N} ET_{oi,j,y,d} \tag{7}$$

The climatology was derived exclusively from the 2014–2017 training period ($N = 4$). Observations on February 29, 2016 were pooled with those on February 28, and the resulting daily climatology was smoothed using a 31-day centered moving average applied cyclically across the calendar year to reduce sampling noise from the short record. This baseline represents the mean seasonal and spatial patterns of ET_o but does not account for year-specific weather anomalies, thereby providing a reference for evaluating whether the CNN-LSTM produced predictive skill beyond the average seasonal cycle.

2.2.2.3. Persistence-based forecasting. Anomaly persistence forecasts (Murphy, 1992) were used as a second baseline. Unlike climatology, this forecast depends on the conditions at the time of forecast issue. For each grid cell (i,j), the ET_o anomaly on the forecast issue day, defined as the departure of the observed ET_o from the climatological forecast (Eq. 7), was assumed to persist over the forecast horizon and was added to the climatological forecast of each target day:

$$\widehat{ET}^{pers}_{o,i,j,d+h} = \widehat{ET}^{clim}_{o,i,j,d+h} + \left[ET_{o,i,j,d} - \widehat{ET}^{clim}_{o,i,j,d}\right], \quad h = 1, \quad \ldots, \quad 56 \tag{8}$$

where d denotes the day of year of the forecast issue day (the last day of the input sequence) in 2018, $d+h$ denotes the day of year of the target day at lead h, and $ET_{o,i,j,d}$ is the observed ET_o on the issue day. This baseline combines the seasonal cycle with the persistence of year-specific conditions at forecast issue. It differs from the raw persistence baseline in the CNN-LSTM residual-learning formulation, which holds the last observed ET_o constant across all lead times without accounting for the seasonal cycle.

2.2.3. Empirical forecasting of actual evapotranspiration and actual crop coefficient

To complement the deep-learning ET_o forecasts, we implemented one empirical forecasting method, termed Lead-Ratio Updating (LRU), to generate forecasts of daily $ET^{O}_{c\ act}$ (Fig. 2a and c) and $K_{c\ act}$ (Fig. 2b and d) at the EC flux tower sites. The LRU method assumes that the relative evolution of crop water use over future lead times can be approximated by the corresponding lead-time-dependent seasonal pattern observed in the previous year.

For $ET^{O}_{c\ act}$ forecasting, we first constructed a smoothed $ET^{O}_{c\ act}$ series using a moving-average window (default: 30 days) to reduce day-to-day noise in the ratio. Based on the 2017 smoothed $ET^{O}_{c\ act}$, LRU maps each anchor day t to a set of multiplicative ratios $r_h(t+h)$ for each forecast lead h (Equation 9). Forecasts of $ET^{O}_{c\ act}$ in 2018 were then generated by applying these lead-specific ratios to the baseline $ET^{O}_{c\ act}(t)$ available on the anchor day (Eq. 10).

$$r_h(t+h) = \frac{ET^{O}_{c\ act}(t+h)}{ET^{O}_{c\ act}(t)}, \quad h = 1, \quad \ldots, \quad H. \tag{9}$$

$$\widehat{ET}^{O}_{c\ act}(t+h) = ET^{O}_{c\ act}(t) \bullet r_h(t+h) \tag{10}$$

Daily $K_{c\ act}$ was derived by pairing the smoothed $ET^{O}_{c\ act}$ series (Fig. 2a and Fig. 2c) with the ET^{PC}_{o} (Eq. 11). The same LRU framework was then applied to the 2017 $K_{c\ act}$ series, mapping each anchor day t to a set of multiplicative ratios $r'_h(t+h)$ for each forecast lead h (Eq. 12). Forecasts of $K_{c\ act}$ in 2018 were then generated by applying these lead-specific ratios to the baseline $K_{c\ act}(t)$ available on the anchor day (Eq. 13).

$$K_{c\ act}(t) = \frac{ET^{O}_{c\ act}(t)}{ET^{PC}_{o}(t)} \tag{11}$$

$$r'_h(t+h) = \frac{K_{c\ act}(t+h)}{K_{c\ act}(t)}, \quad h = 1, \quad \ldots, \quad H. \tag{12}$$

$$\widehat{K_{c\ act}}(t+h) = K_{c\ act}(t) \bullet r'_h(t+h) \tag{13}$$

2.2.4. Model performance evaluation metrics

Model performance was quantified using the coefficient of determination (R^2), root mean square error (RMSE), and percent bias (PBIAS), representing explained variance, error magnitude, and systematic bias, respectively. These metrics were computed across all experiments to evaluate agreement between daily ET_o forecasts and observations. Comprehensive implementation details, including diagnostic tools such as 1:1 scatter plots and residual analysis, are available in the accompanying open-source Python repository (Gao, 2026b).

For the multi-lead ET_o forecasts, performance was additionally evaluated relative to the climatology (Eq. 7) and anomaly persistence (Eq. 8) baselines, which were verified on exactly the same grid cells, issue days, and target days as the CNN-LSTM. At each lead h, the improvement of the CNN-LSTM over each baseline was quantified using the mean squared error skill score (SS; Murphy, 1988):

$$SS(h) = 1 - \frac{\sum (\widehat{ET}_o^{CNN-LSTM} - ET_o)^2}{\sum (\widehat{ET}_o^{ref} - ET_o)^2} \quad (14)$$

where $\widehat{ET}_o^{CNN-LSTM}$ and $\widehat{ET}_o^{ref}$ denote the CNN-LSTM and baseline forecasts, respectively, ET_o is the observed value, and the summations are over all grid cells and verification cases at lead h. SS was computed separately with climatology (Eq. 7) and anomaly persistence (Eq. 8) as the baseline, hereafter denoted SS^{clim} and SS^{pers}, respectively. SS is analogous to the Nash-Sutcliffe efficiency but uses a baseline forecast rather than the observed mean as the reference. $SS > 0$ indicates that the CNN-LSTM outperformed the baseline. The limit of the forecasting skill was defined as the first lead at which the lower bound of the 95% confidence interval of SS reached zero.

To assess forecasting information beyond the seasonal cycle, the $\widehat{ET}_o^{CNN-LSTM}$ and ET_o were converted to anomalies, denoted by primes, by subtracting the climatology of the target day, and the anomaly correlation coefficient (ACC; Wilks, 2019) was computed at each lead h:

$$ACC(h) = \frac{\sum (\widehat{ET}_o^{'\ CNN-LSTM} - \overline{\widehat{ET}_o^{'\ CNN-LSTM}})(ET_o' - \overline{ET_o'})}{\sqrt{\sum (\widehat{ET}_o^{'\ CNN-LSTM} - \overline{\widehat{ET}_o^{'\ CNN-LSTM}})^2 \sum (ET_o' - \overline{ET_o'})^2}} \quad (15)$$

where the overbars denote means, and the summations are over all grid cells and verification cases at lead h. ACC ranges from -1 to 1, and values above 0.5 are often used to indicate useful skill in extended-range forecasting. Confidence intervals for SS and ACC were estimated using a moving-block bootstrap (Künsch and Hans, 1989) over issue days (block length = 7 days; 1000 resamples) to account for serial correlation among forecasts issued on consecutive days. In addition, PBIAS was computed for the raw persistence baseline of the residual-learning formulation to diagnose the source of bias at longer leads.

3. Results

3.1. Continuous wavelet transform analysis

The dominant time periods across the AOI changed as P_{max} increased. The black line (Fig. 4a) shows changes in the number of unique dominant time periods as longer periods were included in the analysis. Fig. 4b shows their spatial distribution at the selected (P_{max}), while Fig. A5 provides maps for 12 additional P_{max} settings. Together, these results show that both the number and spatial distribution of dominant time periods depended on the upper bound of the analyzed period range.

The 26- and 93-day periods each remained stable within the MOI over a specific range of P_{max} values (Fig. 4a). The 26-day period all MOI pixels when the analysis was limited to shorter periods (when P_{max} is between 28 days and 56 days) as shown by the blue line in Fig. 4a. The 93-day period then occupied approximately 100% of the MOI from $P_{max} = 98$ to $P_{max} = 126$ days, as shown by the orange line. No other dominant period showed comparable spatial persistence across multiple consecutive P_{max} settings.

Instead of 26-day period, the 93-day period was selected because once the representative dominant timescale for the MOI. This is considering that once P_{max} was expanded to include both periods, the 93-day period replaced the 26-day period and remained dominant throughout the MOI from 98 days to 126 days. Importantly, this selection was supported by consistently higher global wavelet power at approximately 93-day scale than at the 26-day scale across MOI pixels (Fig. A2). The 93-day period was therefore used to inform the candidate input lengths for the subsequent CNN-LSTM experiments.

3.2. Reference evapotranspiration forecasting performance within the modeling area of interest

The training, validation, and testing subsets covered similar ET_o ranges but differed modestly in their distributions. Values in all three subsets ranged from approximately 0 to around 10 mm•day^{-1} (Fig. A4). However, the testing subset showed a different density distribution at lower ET_o values, particularly between around 1 to 3 mm • day^{-1}.

Input sequence length affected CNN-LSTM performance, with $L_{in} = 60$ days providing the best overall results among the tested configurations (Fig. 5). The dominant ET_o period identified by CWT (93 days) did not directly correspond to the best input sequence length for CNN-LSTM forecasting. Increasing L_{in} from 60 to 93 days significantly reduced performance across the 56 lead days, with mean R^2 decreasing

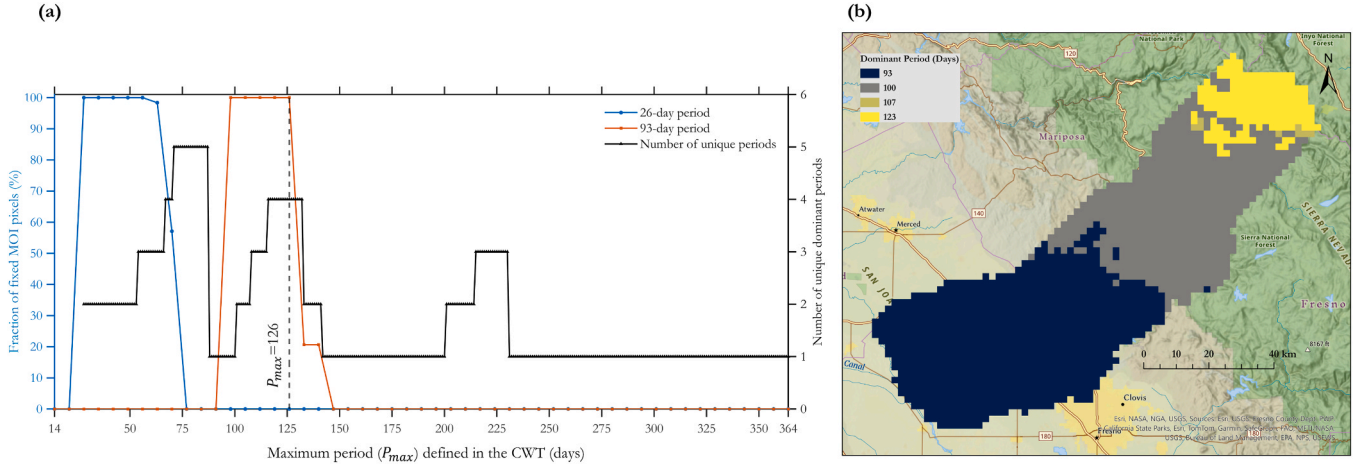

Fig. 4. Sensitivity of CWT-derived dominant periods to the maximum analyzed period, P_{max}. (a) Number of unique dominant periods across the AOI (black line, right axis) and fractions of MOI pixels classified as having 26-day (blue line) and 93-day (orange line) dominant periods (left axis) as P_{max} increased from 14 to 364 days in 7-day increments. The vertical dashed line marks the selected P_{max} of 126 days. (b) Spatial distribution of the four dominant periods (93, 100, 107, and 123 days) obtained using the selected 126-day configuration as the P_{max}. Maps for 12 additional P_{max} configurations are provided in Fig. A 5.

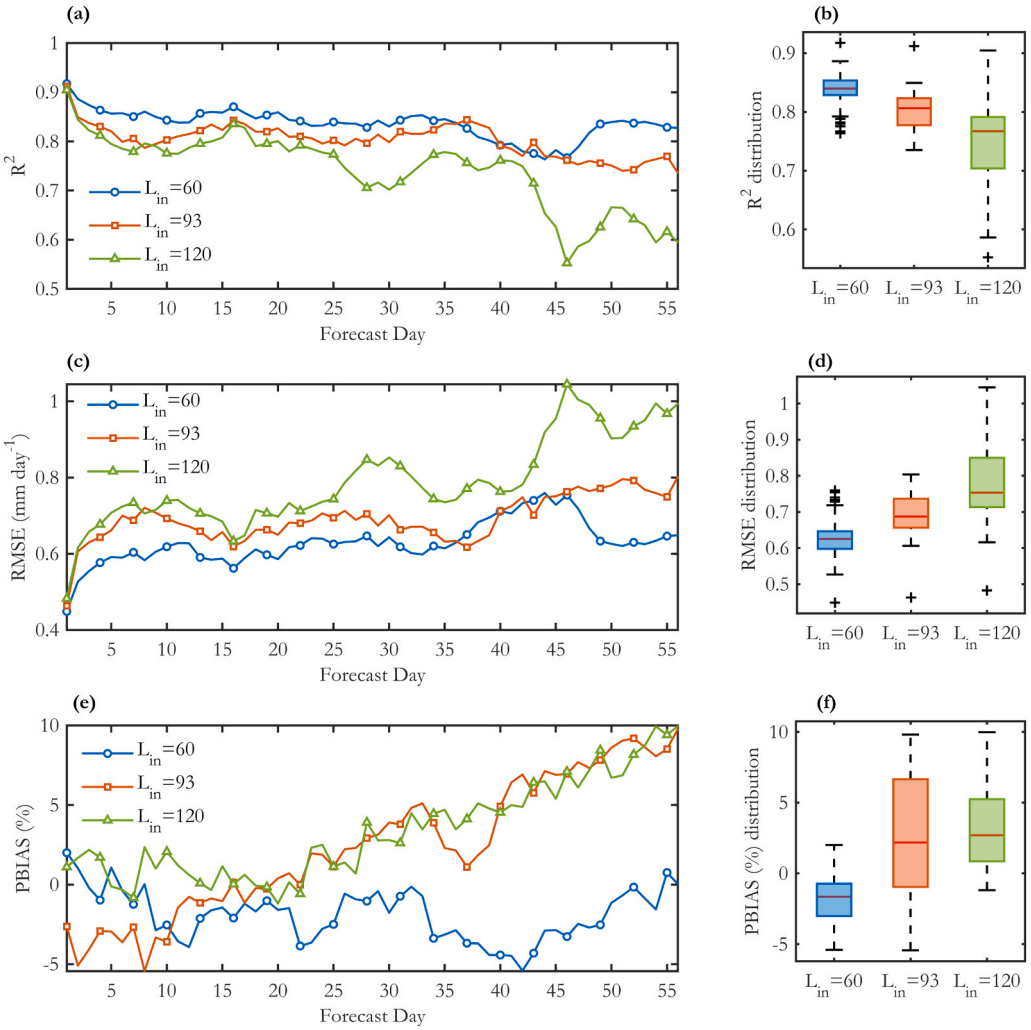

Fig. 5. CNN-LSTM ET_o forecasting performance across lead days 1–56 for three different input sequence lengths (L_{in} = 60, 93, and 120 days). The left panels show R^2, RMSE, and PBIAS as functions of forecast lead time (a, c, and e), while the right panels show their distributions across the 56 lead days (b, d, and f). The horizontal dashed lines indicate the overall climatological-reference values for R^2, RMSE, and PBIAS. All metrics were computed only for forecasts whose target days fell between August 1st and October 31st, 2018. Positive PBIAS indicates overestimation.

from 0.84±0.03to 0.80±0.03 and mean RMSE increasing from 0.63±0.06 mm•day^{-1} to 0.69±0.06 mm•day^{-1}. A further increase to L_{in}= 120 days reduced mean R^2 to 0.74±0.08 and increased mean RMSE to 0.79±0.12 mm•day^{-1}, and both changes were statistically significant (α=0.05).

PBIAS differed among input lengths, particularly at longer leads (Fig. 5e, f). For L_{in}= 60 days, PBIAS remained between approximately − 5% and 2% across all leads and was negative at most leads (median≈ − 1.8%), indicating a slight tendency toward underestimation. For L_{in}= 93 and 120 days, PBIAS generally increased with lead time. During approximately the first three weeks, PBIAS for L_{in}= 93 days rose from about − 5% to near 0%, whereas that for L_{in}= 120 days remained close to 0%. Beyond this period, both configurations showed a similar, continued increase, reaching about 10% by lead day 56.

Overall, the CNN-LSTM performed between the two benchmarks (Fig. 6a-c). Across nearly all leads, SS^{clim} was negative, whereas SS^{pers} was positive or close to zero, indicating that the CNN-LSTM was generally less accurate than climatology but comparable to or more accurate than anomaly persistence. This pattern was most evident for L_{in}= 60 days, whereas for L_{in}= 120 days SS^{pers} became negative from lead day 31 onward.

Relative to climatology, SS^{clim} was positive only on lead day 1 under all three input lengths (SS^{clim} =0.2–0.3; Fig. 6a-c). However, the 95% confidence intervals included zero in all cases. From lead day 2 onward, SS^{clim} was negative for all configurations, indicating larger errors than the climatological forecast, and the deficit increased with input length. The interquartile range of SS^{clim} across the 56 lead days was − 0.55 to − 0.32 for L_{in}= 60 days, − 1.02 to − 0.60 for L_{in}= 93 days, and − 1.68 to − 0.89 for L_{in}= 120 days. The minimum SS^{clim} was about − 1.1 on lead day 44 for L_{in}= 60 days, − 1.4 on lead day 56 for L_{in}= 93 days, and − 3.1 on lead day 46 for L_{in}= 120 days. Thus, none of the tested configurations showed statistically significant skill relative to climatology at any lead.

After removing the climatology, *ACC* was approximately 0.57 on lead day 1 for all three input lengths (95% confidence interval: 0.36–0.67; Fig. 6d-f), within the range (0.5–0.6) commonly regarded as indicating useful skill (Albers and Newman, 2019). *ACC* then decreased rapidly to 0.30–0.33 on lead day 2 and approached zero within about one week. The 95% confidence intervals excluded zero on lead days 1 and 2 for all configurations but included zero at nearly all subsequent leads. The exceptions were brief periods of significant positive *ACC* around lead days 16–17 (maximum of 0.31 for L_{in}=120 days) and significant negative *ACC* around lead days 44–48 for all configurations, where *ACC* reached minima of − 0.30, − 0.25, and − 0.36, for L_{in}= 60, 93, and 120 days, respectively.

The CNN-LSTM corrected most of the systematic bias inherent in its

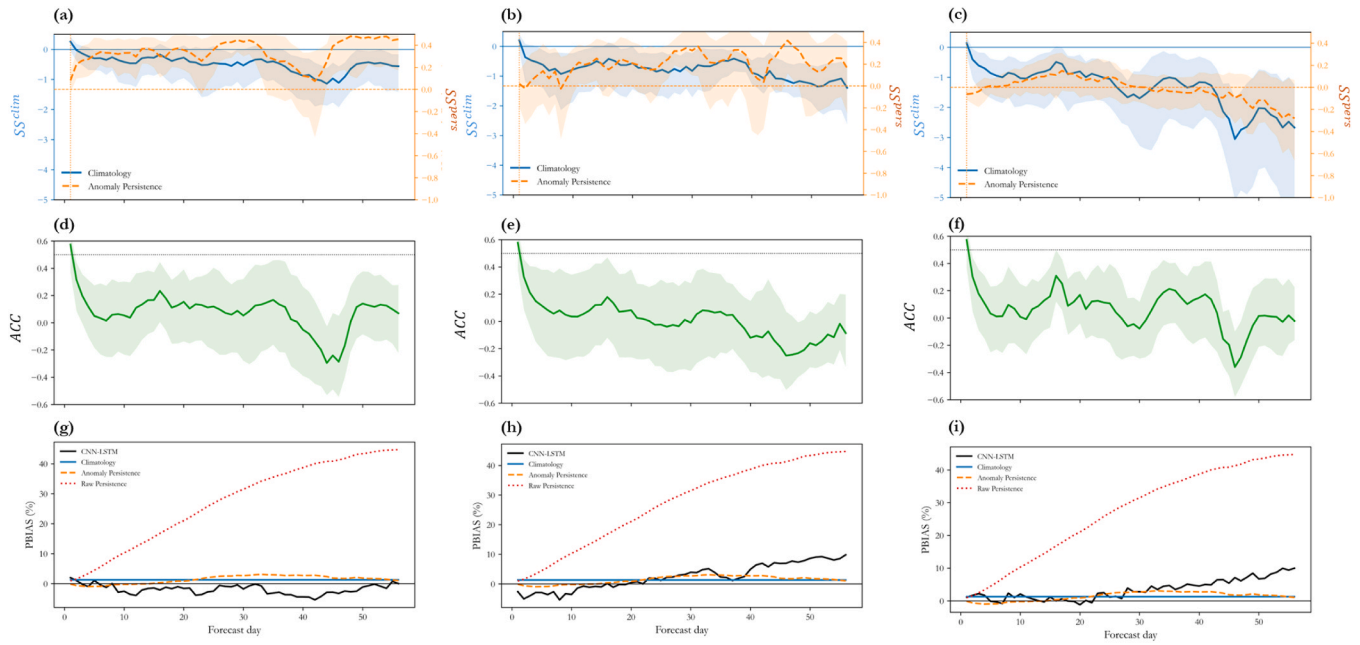

Fig. 6. Benchmark-relative skill of the CNN-LSTM ET_o forecasts across lead days 1–56 for input sequence lengths of L_{in}= 60 days (left column), L_{in}= 93 days (middle column), and L_{in}= 120 days (right column). (a-c) Mean squared error skill score (*SS*) relative to climatology and anomaly persistence. Horizontal lines mark *SS*= 0 on the corresponding axes, and vertical dotted lines indicate the first lead at which the lower bound of the 95% confidence interval of *SS* reached zero. (d-f) Anomaly correlation coefficient (*ACC*) between the CNN-LSTM forecast and observed ET_o anomalies after removing the climatology; the dotted line marks *ACC* = 0.5. (g-i) Percent bias (PBIAS) of the CNN-LSTM, climatology, anomaly persistence, and raw persistence forecasts, where raw persistence denotes the baseline of the residual-learning formulation in which the last observed ET_o was held constant across all leads; positive values indicate overestimation. Shaded areas denote 95% confidence intervals. All forecasts were verified on the same time period: from August 1st to October 31st, 2018.

raw persistence baseline (Fig. 6g-i). The raw persistence baseline of the residual-learning formulation showed a PBIAS that increased steadily with lead time, from near zero on lead day 1 to about 45% on lead day 56 (Fig. 6g-i). In contrast, the PBIAS of the CNN-LSTM remained within approximately − 5%to 2% across all leads for L_{in}= 60 days, whereas a residual overestimation remained at longer leads for L_{in}= 93 and 120 days, reaching about 10% by lead day 56. The PBIAS of the climatology and anomaly persistence forecasts remained within approximately 0–3% across all leads.

Taken together, the CNN-LSTM largely corrected the bias of its raw persistence baseline but did not outperform climatology at any lead, and its predictive information beyond the seasonal cycle was limited to the first two lead days. Among the tested configurations, L_{in}= 60 days consistently showed the smallest deficit relative to climatology and the smallest bias.

3.3. Field-scale relationship between gridded and local reference evapotranspiration

ET_o^P was consistently higher than PM-ET_o and H-ET_o at both sites (Fig. 7). Although ET_o^P reproduced the overall temporal decline during the evaluation period, clear magnitude differences persisted at both sites. At RIP720 and RIP760, agreement between ET_o^P and PM-ET_o yielded RMSE values of 0.99 and 1.20 mm•day^{-1}, with corresponding R^2 values of 0.60 and 0.42, respectively. The negative PBIAS values further indicate that PM-ET_o was systematically lower than ET_o^P during the evaluation period. In contrast, ET_o^P showed closer agreement with H-ET_o estimates, with RMSE values of 0.50 and 0.53 mm day^{-1} and R^2 values of 0.90 and 0.89 at RIP720 and RIP760, respectively. Because the PM method was treated as the more physically based site reference, the weaker agreement between ET_o^P and PM-ET_o indicates that local correction or calibration is necessary before applying spatial CIMIS ET_o to field-scale irrigation analyses. Accordingly, after CNN-LSTM forecasting, the linear correction relationship shown in Fig. 7 was applied to the forecasted ET_o before field-scale evaluation and subsequent coupled $ET_{c\ act}$ estimation.

After local correction, field-scale ET_o forecasts agreed closely with ET_o^{PC} at short leads at both vineyard sites, and agreement decreased with increasing lead time (Fig. 8). Through lead day 6, R^2 exceeded 0.90 and RMSE remained below 0.40 mm•day^{-1} at both sites. Through lead day 9, R^2 exceeded 0.80 and RMSE remained below 0.60 mm•day^{-1}. R^2 remained above 0.50 and RMSE below 0.90 mm•day^{-1} through lead days 18 and 16 at RIP720 and RIP760, respectively.

PBIAS also showed a lead-time-dependent pattern, becoming increasingly negative with increasing lead time at both sites (Fig. A6a). This indicates that the field-scale ET_o forecasts increasingly underestimated ET_o^{PC} at longer leads. Overall, agreement between forecasts and ET_o^{PC} was highest within the first 6 days and decreased substantially beyond approximately two weeks.

3.4. Actual evapotranspiration forecasting

3.4.1. Field-scale evaluation of empirical actual evapotranspiration forecasts

Accounting for energy-balance closure substantially changed the evaluated performance of the empirical LRU forecasts, although the results remained strongly site dependent (Fig. 9). At RIP720, evaluation against the energy-closure-corrected EC estimates yielded R^2 values of approximately 0.67–0.81 and RMSE values of approximately 0.56–0.76 mm•day^{-1}, compared with R^2 values near 0–0.19 and RMSE values of approximately 0.98–1.16 mm•day^{-1} when direct EC observations were used. At RIP760, the correction also increased R^2 and reduced RMSE, but R^2 remained negative across all lead days, ranging from approximately −0.9 to −1.6, and RMSE ranging from approximately 1.38–1.62 mm•day^{-1}. PBIAS was also lower when the forecasts were evaluated against the corrected EC estimates at both sites, although it remained substantially higher at RIP760 (Fig. A6b).

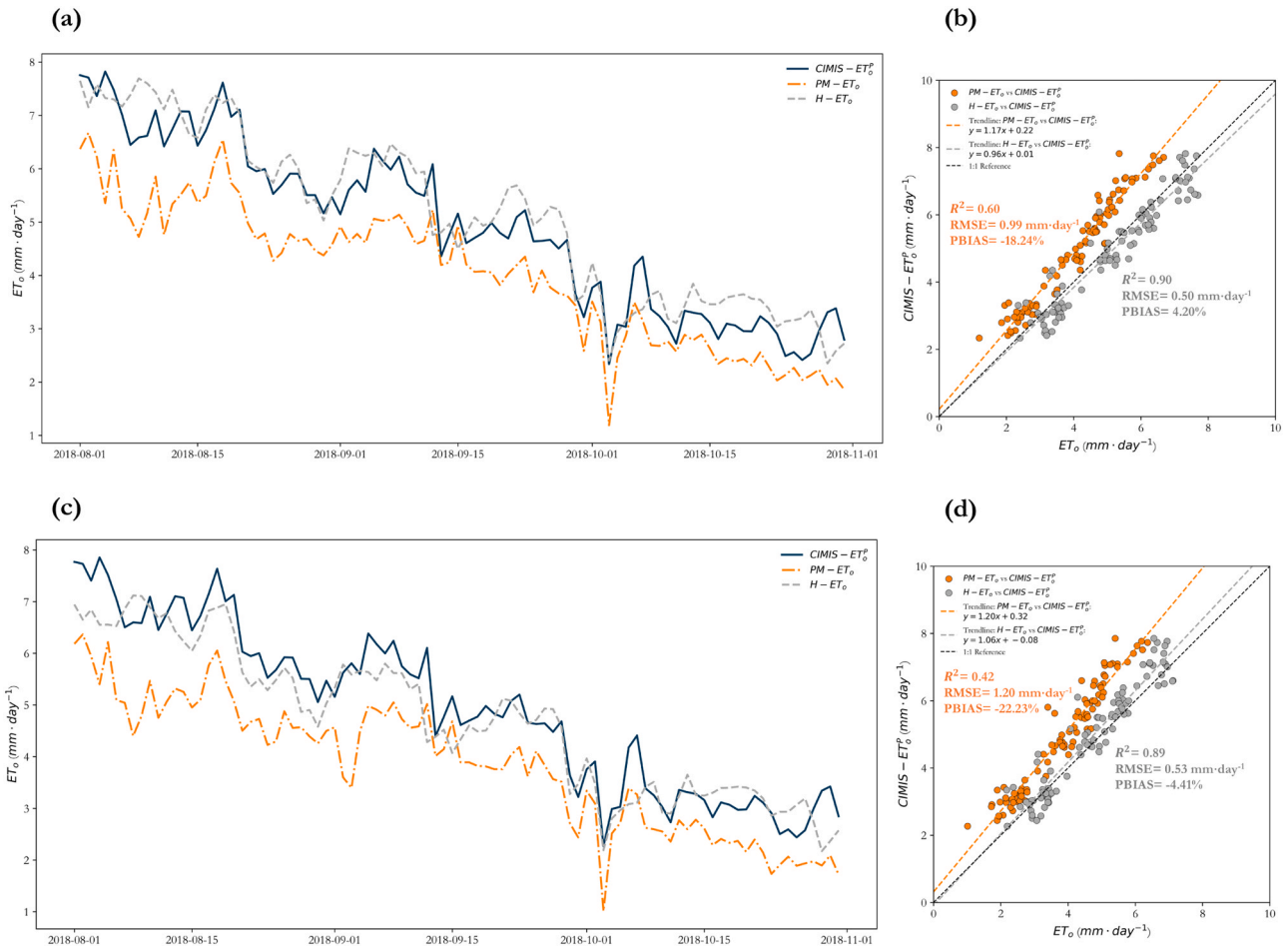

Fig. 7. Field-scale comparison between ET_o^P and estimated ET_o at the EC flux tower sites (PM-ET_o and H-ET_o). Panels (a, b) and (c, d) show results for RIP720 and RIP760, respectively. For each site, the left panel shows daily time series of ET_o^P, PM-ET_o, and H-ET_o during the independent evaluation period from August 1st to October 31st, 2018. The right panel shows corresponding scatter plots comparing ET_o^P with PM-ET_o and H-ET_o. The dashed black line denotes the 1:1 reference line, while the colored dashed trendlines show linear regressions.

3.4.2. *Field-scale evaluation of empirical actual crop coefficient forecasts*

The empirical $K_{c\ act}$ forecasts were site-dependent and non-monotonic across lead times at both sites (Fig. 10), with substantially better overall forecast skill than the corresponding empirical $ET_{c\ act}$ forecasts, but with practical usefulness still largely confined to short lead times. Forecast skill declined during approximately the first two weeks, then partially recovered to a secondary peak near five weeks, before decreasing again later in the forecast horizon. At RIP720, R^2 decreased from 0.37 on lead day 1 to 0.01 on lead day 14, became negative thereafter, briefly returned to positive values between lead days 36 and 39 with a local maximum of 0.02 on lead day 37, and then declined to −0.40 by lead day 56. Over the same period, RMSE generally increased from 0.13 to 0.19 (mm•day^{-1} / mm•day^{-1}), despite some fluctuation during the mid-horizon period. At RIP760, R^2 declined from 0.53 on lead day 1to 0.05 on lead day 17, increased to 0.35 on lead day 37, and then decreased to −0.11 by lead day 56. RMSE followed a complementary pattern, increasing from 0.14 to 0.20 (mm•day^{-1} / mm•day^{-1}) by lead day 17, decreasing to 0.17 (mm•day^{-1} / mm•day^{-1}) on lead day 37, and then increasing again to 0.22 (mm•day^{-1} / mm•day^{-1}) by lead day 56. The corresponding PBIAS results showed a gradual increase with forecast lead time at both sites, but the magnitude remained relatively small, generally within approximately 0–5% (Fig. A6c). The positive PBIAS values indicate a slight tendency to overestimate $K_{c\ act}$, particularly at longer lead times. Overall, these results suggest that the main limitation of empirical $K_{c\ act}$ forecasting was not large systematic bias, but weak explanatory skill and site-dependent uncertainty, which can subsequently constrain coupled $ET_{c\ act}$ estimation.

3.4.3. *Field-scale evaluation of coupled actual evapotranspiration forecasts*

The evaluated skill of the coupled $ET_{c\ act}$ forecasts differed substantially depending on whether direct or energy-closure-corrected EC estimates were used as the observational reference, and the magnitude of this difference varied between sites (Fig. 11 and Fig. A7).

At RIP720, forecast performance followed a similar three-phase pattern under both EC references, with consistent timing: performance improved during approximately the first three weeks, deteriorated until lead day 42 (week 6), and partially recovered toward lead day 56 (Fig. 11b). When evaluated against direct EC observations, R^2 increased during approximately the first three weeks, then generally decreased, despite some fluctuations, to 0.70 on lead day 42, before increasing again to 0.74 on lead day 56. RMSE showed the opposite pattern, decreasing from 0.71 mm•day^{-1} on lead day 1to 0.65 mm•day^{-1} on lead day 21, then generally increasing to 0.79 mm•day^{-1} on lead day 42, before decreasing to 0.74 mm•day^{-1} on lead day 56. When evaluated against the energy-closure-corrected EC estimates, performance was lower throughout the forecast horizon, although the timing of these changes was consistent. R^2 increased from 0.63 on lead day 1to 0.68 at approximately three weeks, decreased to 0.47 on lead day 42, and then increased slightly to 0.49 on lead day 56. RMSE remained relatively stable at approximately 0.93 mm•day^{-1} during the first three weeks, increased to 1.20 mm•day^{-1} by lead day 42, and remained at approximately this level thereafter.

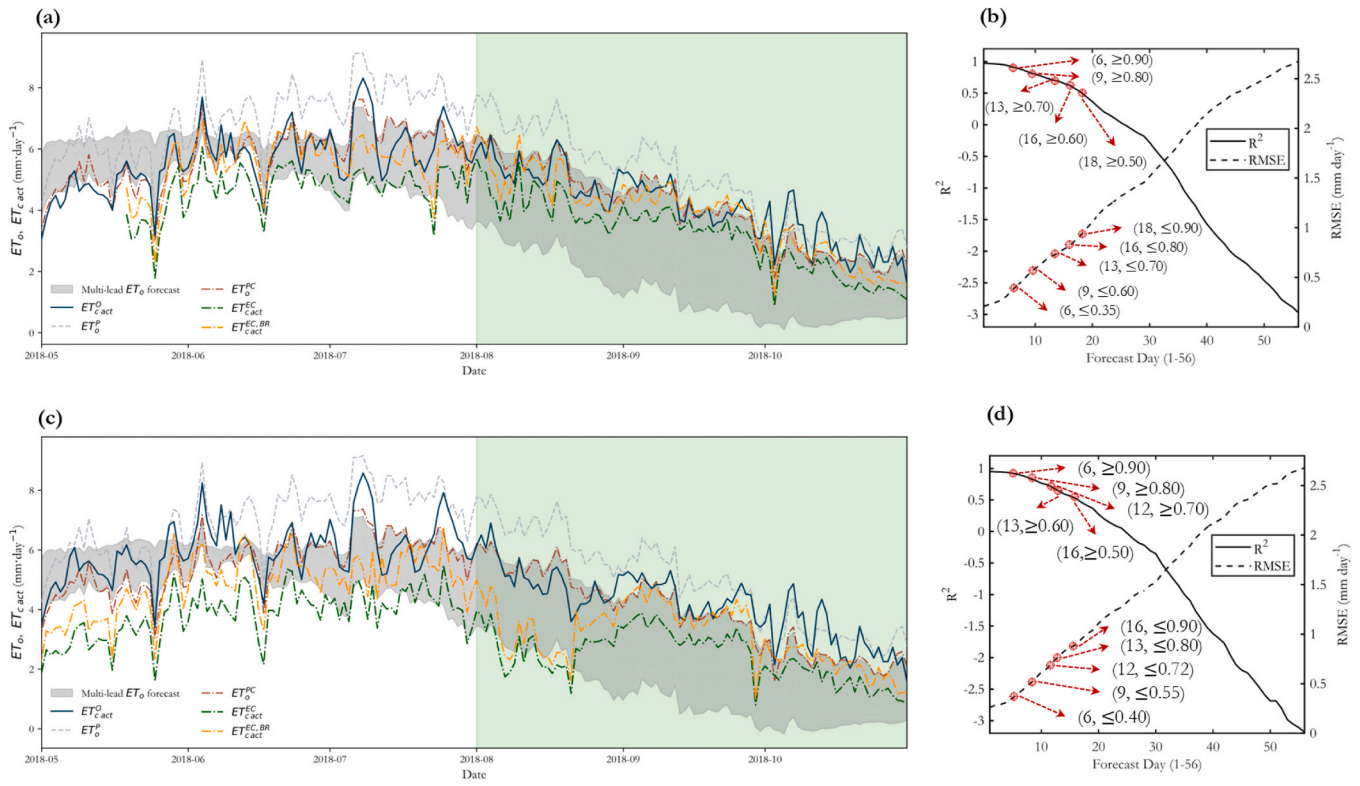

Fig. 8. Evaluation of field-scale ET_o^{PC} forecasting. Panels (a) and (c) show time-series plots of multi-lead ET_o^{PC} forecasts together with ET_o^{P} and ET_o^{PC} at RIP720 and RIP760, respectively. $ET_{c\ act}^{O}$ and $ET_{c\ act}^{EC}$ are shown as background reference series. The light green shaded area indicates the independent evaluation period from August to October 2018. Panels (b) and (d) show forecast performance across lead days 1–56, summarized by R^2 and RMSE (mm•day^{-1}) for the August-October 2018 period. The corresponding PBIAS results are presented in Fig. A 6a.

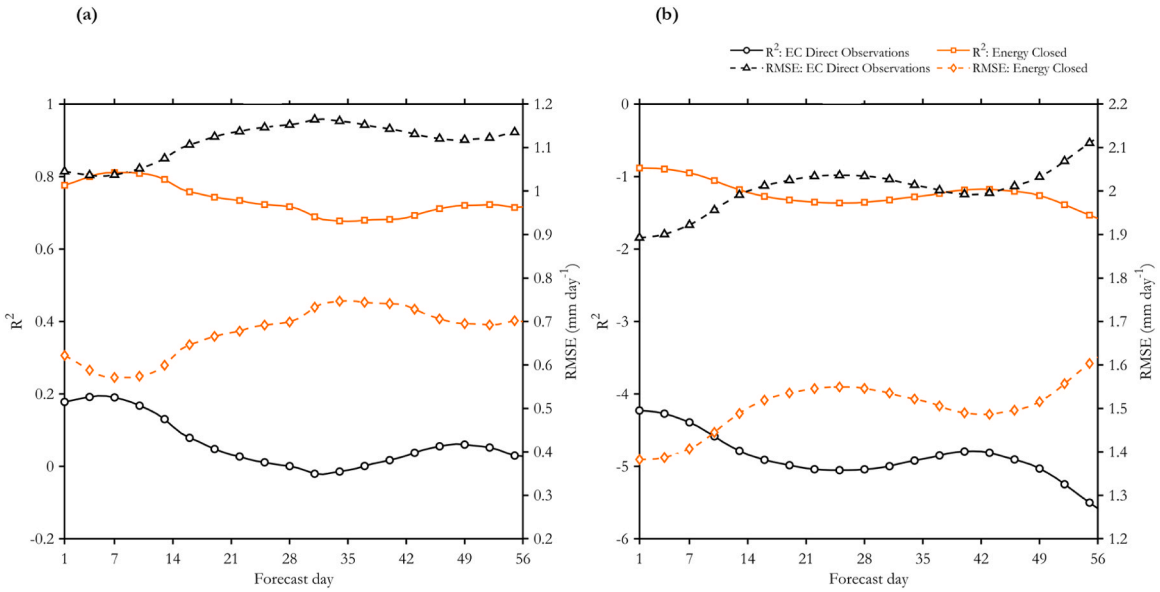

Fig. 9. Performance of the empirical LRU forecasts of $ET_{c\ act}$ (i.e., $ET_{c\ act}^{LRU}$) at RIP720 (a) and RIP760 (b) across lead days 1–56 during the August–October 2018 evaluation period. Black lines show evaluation against direct EC observations ($ET_{c\ act}^{EC}$), and orange lines show evaluation against Bowen-ratio corrected EC estimates ($ET_{c\ act}^{EC.BR}$). The corresponding PBIAS results are presented in Fig. A 6b.

At RIP760, by contrast, forecast performance was higher when evaluated against the energy-closure-corrected EC estimates, and both references showed a three-phase pattern with consistent timing: performance deteriorated until the end of week 4 (lead day 29), improved until approximately lead days 44–47 (weeks 6–7), and declined slightly toward lead day 56 (Fig. 11d). When evaluated against direct EC observations, R^2 remained negative during approximately the first five weeks and, despite some fluctuations, generally decreased to its minimum of − 0.10 on lead day 29. It then increased to its maximum of 0.09 on lead day 44, before decreasing to 0.02 on lead day 56; overall, R^2

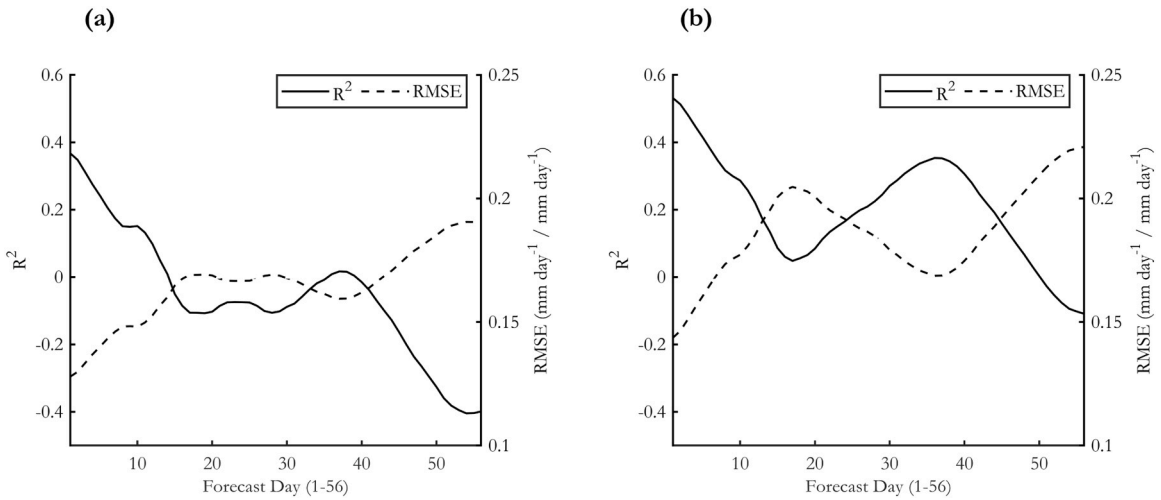

Fig. 10. Forecast performance of the empirical LRU method for $K_{c\ act}$ (i.e., K_{cact}^{LRU}) at RIP720 (a) and RIP760 (b) across lead days 1–56, expressed as R^2 and RMSE (mm•day^{-1}/mm•day^{-1}) during the August – October 2018 evaluation period. The corresponding PBIAS results are presented in Fig. A6c.

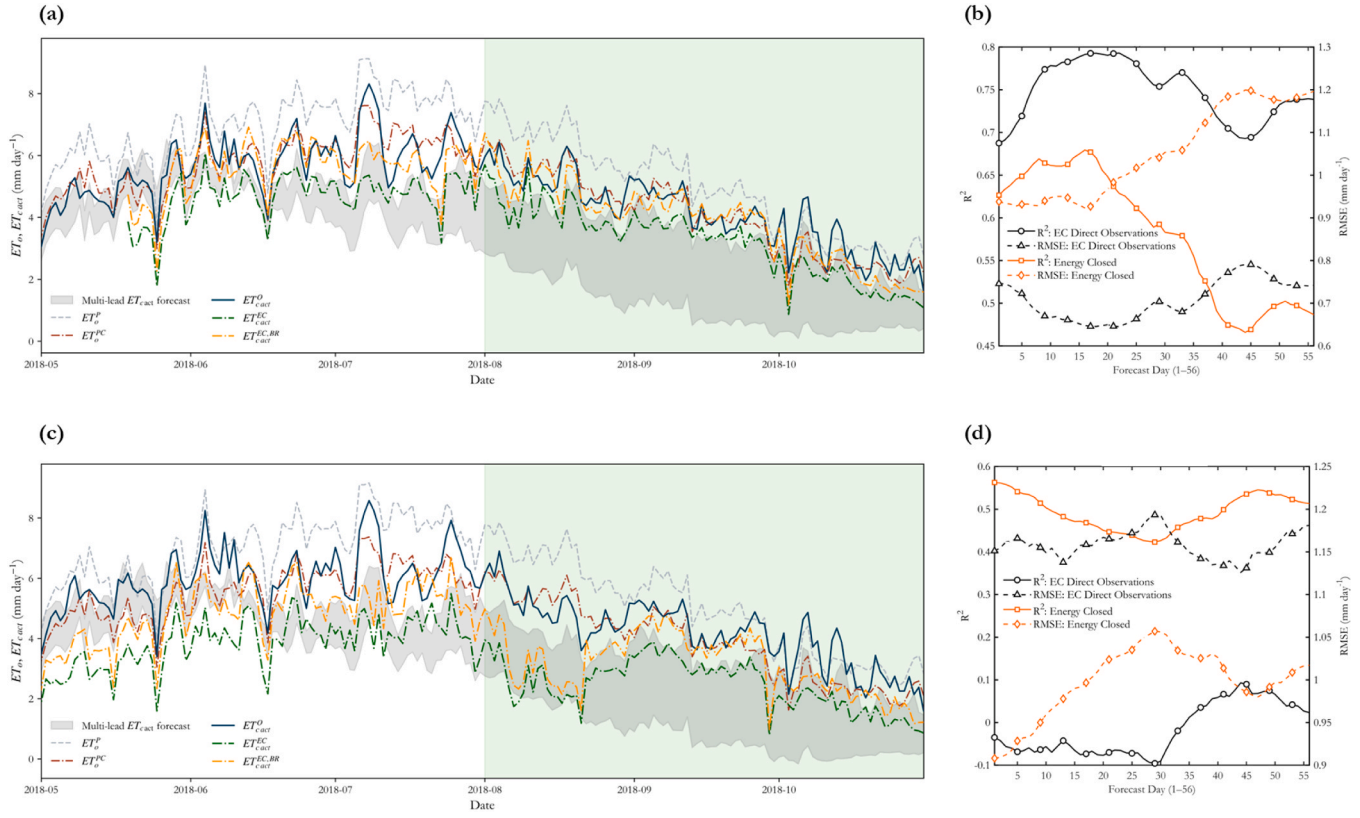

Fig. 11. Field-scale evaluation of coupled $ET_{c\ act}$ forecasts generated by combining ET_o^{PC} forecasts with $K_{c\ act}^{LRU}$ at RIP720 and RIP760. Panels (a) and (c) show the coupled forecasts together with $ET_{c\ act}^{EC}$, $ET_{c\ act}^{EC,BR}$, $ET_{c\ act}^{O}$, ET_o^P, and ET_o^{PC} at RIP720 and RIP760, respectively. The light green shading indicates the independent evaluation period from August through October 2018. Panels (b) and (d) show R^2 and RMSE across lead days 1–56 at RIP720 and RIP760, respectively. Black lines indicate evaluation against $ET_{c\ act}^{EC}$, while orange lines indicate evaluation against $ET_{c\ act}^{EC,BR}$.

varied within a narrow range of approximately 0.20. RMSE showed the opposite pattern, generally increasing to its maximum of 1.20 mm•day^{-1} on lead day 29, decreasing to its minimum of 1.13 mm•day^{-1} on lead day 44, and then increasing again to 1.18 mm•day^{-1} on lead day 56. When evaluated against the energy-closure-corrected EC estimates, the changes within each phase were more monotonic. R^2 decreased from 0.56 on lead day 1to 0.42 on lead day 29, increased to 0.55 on lead day 47, and then decreased to 0.51 on lead day 56. Correspondingly, RMSE increased from 0.91 mm•day^{-1} on lead day 1to 1.06 mm•day^{-1} on lead day 29, decreased to 0.98 mm•day^{-1} on lead day 47, and increased to 1.02 mm•day^{-1} on lead day 56.

The PBIAS results further showed that the choice of observational reference affected both the magnitude and the sign of the evaluated forecast bias (Fig. A8). At RIP720, PBIAS against direct EC observations increased from − 7.4% on lead day 1 to near zero by lead day 35 and

remained slightly positive thereafter, whereas PBIAS against the energy-closure-corrected estimates was positive throughout the forecast horizon and increased from 10.7% to 19.0%. At RIP760, PBIAS against direct EC observations was consistently negative, ranging from − 27.4% to − 20.5%, whereas PBIAS against the energy-closure-corrected estimates was close to zero, being slightly negative during the first 8 days and positive thereafter. Thus, the reference that yielded the smaller bias differed between sites: direct EC observations at RIP720 and energy-closure-corrected estimates at RIP760.

4. Discussion

4.1. Relationship between the dominant timescale and input sequence length

Continuous wavelet transform analysis was useful for identifying a characteristic timescale in ET_o variability that could inform the design of CNN-LSTM input-sequence experiments. In this study, the dominant timescale of 93 days identified for the southwest Madera area represents a stable sub-seasonal feature in regional ET_o during 2014–2017, rather than a directly verified physical process. Its spatial localization in the more intensively irrigated portion of the county suggests that it may reflect irrigation-related land-atmosphere interactions or other spatially varying surface controls, consistent with prior evidence that irrigation can modify local atmospheric conditions over cultivated areas (Sorooshian et al., 2011). This spatially varying timescale is consistent with broader findings that environmental interactions often exhibit strong spatial structure and driver dependence (Yang et al., 2024). At the same time, a period of approximately 93 days is close to one quarter of the annual cycle, so part of this signal may also reflect the non-sinusoidal shape of the seasonal cycle in ET_o.

However, the dominant timescale did not translate into the best-performing input sequence length. Among the three tested configurations, L_{in}= 60 days produced the highest R^2, the lowest RMSE, and the smallest bias, whereas L_{in}= 93 and 120 days performed worse, particularly at longer leads, where they also developed an overestimation of about 10% (Fig. 5 and Fig. 6). This result can be understood by distinguishing the timescale that dominates ET_o variance from the timescale over which ET_o is predictable. The variance of ET_o is dominated by the seasonal cycle, which is largely deterministic and already captured by climatology; the predictable information beyond this cycle resides in ET_o anomalies, whose memory was short. This is evident from the fact that anomaly persistence was less accurate than climatology and that *ACC* decreased to near zero within about one week (Fig. 6). Consequently, extending the input window to span the dominant 93-day period, or beyond, added little information relevant to forecasting departures from the seasonal cycle, while increasing the difficulty of learning from longer sequences and possibly encouraging the model to fit season-specific features of the training years. Thus, the CWT-derived timescale is better interpreted as a diagnostic of the structure of ET_o variability than as a guide for selecting the input sequence length, and spectral or wavelet analyses used for this purpose should distinguish variance-dominant timescales from predictability timescales.

Because only ET_o was used as model input, the physical basis of the 93-day timescale cannot be determined conclusively. Applying the CWT to ET_o anomalies, with the seasonal cycle removed, would clarify whether this timescale persists beyond the seasonal cycle or whether any sub-seasonal periodicity could more effectively inform input design. Given the short memory of ET_o anomalies, input sequences shorter than 60 days also merit evaluation. In addition, future work could compare direct forecasting based on historical ET_o with an indirect approach that forecasts the meteorological drivers used to calculate ET_o, following Torres et al. (2011), who demonstrated the potential of the indirect approach to improve ET_o forecasts, particularly at longer lead times. Such an approach would also be consistent with recent recommendations that AI/ML methods support, rather than replace, the standardized physical calculation of ET_o (Kukal et al., 2026). Evaluating the dominant timescales of individual meteorological variables, vegetation dynamics, and irrigation-related indicators or records would further clarify how these drivers collectively contribute to the 93-day signal.

In addition, because the dominant timescale was derived from the gridded CIMIS ET_o product (Hart et al., 2009), uncertainty in this product may have affected both the identified timescale and, consequently, the candidate input sequence lengths derived from it; the implications of this product uncertainty for forecast evaluation are discussed in Section 4.2.

4.2. Scale-dependent performance and limitations of reference evapotranspiration forecasting

At the regional scale, the CNN-LSTM reproduced the seasonal evolution of ET_o with relatively high accuracy throughout the 56-day forecast horizon (Fig. 5). Field-scale agreement was lower after local correction was applied to represent site-specific meteorological conditions (Fig. 8). However, these accuracy metrics should be interpreted together with the benchmark evaluation. Because the CNN-LSTM did not significantly outperform climatology at any lead (Fig. 6), its high R2 at the regional scale largely reflects the seasonal cycle shared by forecasts and observations rather than year-specific predictive information. Thus, for water-resources applications at either scale, the forecasts provide limited additional value beyond climatology at extended leads, and their usefulness should be assessed relative to such benchmarks rather than from accuracy metrics alone.

The benchmark results also help explain the limited skill of the CNN-LSTM. The model corrected most of the systematic bias of its raw persistence baseline (Fig. 6g-i), indicating that it successfully learned the seasonal decline in ET_o during the testing period. Its skill in predicting ET_o anomalies, however, was confined to the first two lead days, and *ACC* remained below the range commonly regarded as indicating useful skill thereafter (Fig. 6d-f). Under such conditions, forecasts that retain anomalies with only weak correlation to observations are expected to be less accurate than climatology, which predicts no anomaly; for an unbiased forecast with realistic variance, the mean squared error skill score relative to climatology is positive only when the anomaly correlation exceeds 0.5 (Murphy and Epstein, 1989). In addition, anomaly persistence was itself less accurate than climatology, indicating that ET_o anomalies decorrelated within a few days. Together, these results suggest that, with historical ET_o as the only input, the predictable information available beyond the seasonal cycle is largely exhausted within the first week. Improving extended-range ET_o forecasts is therefore likely to require additional predictors, such as sub-seasonal weather forecasts, or model formulations that use climatology rather than raw persistence as the baseline for residual learning.

The uncertainty of the ET_o product itself further complicates the interpretation of forecast skill. The spatial CIMIS ET_o used in this study is not a direct observation but a modeled product, generated at 2-km resolution by combining satellite-derived parameters (e.g., solar radiation) from GOES imagery with spatially interpolated station measurements of other meteorological variables (Hart et al., 2009). Errors in these components, such as those arising from cloud contamination in the satellite retrievals or from interpolation across sparse station networks, propagate into the gridded ET_o and are likely to vary in space and time. Because the same product served as the model input, the training target, and the reference for evaluating both the CNN-LSTM and the climatological benchmark, errors inherent in the product cannot be separated from forecast errors, and the evaluated skill reflects agreement with the product rather than with the ET_o calculated based on local meteorological situations. The systematic differences between CIMIS ET_o and PM-ET_o at the vineyard sites (Fig. 7) indicate that such errors can be substantial at the field scale. Without a quantitative characterization of the product uncertainty, the uncertainty of the resulting forecasts cannot be fully assessed. Further studies should therefore evaluate the gridded

ET_o against independent station-based PM-ET_o across a broader range of locations and conditions, and consider propagating the product uncertainty into probabilistic forecasts.

A key limitation of the present study is the relatively short temporal span used for model development and evaluation. We selected the 2014–2018 ET_o products from CIMIS based on the available site observations and partitioned them into training (2014–2017), validation (2018–05–01–2018–07–31), and testing (2018–08–01–2018–10–31). Consequently, the model was trained on only four growing seasons, and independent testing was limited to the latter part of a growing season. In addition, the linear correction between CIMIS ET_o and PM-ET_o was calibrated and evaluated over the same August-October 2018 period, which may introduce in-sample bias and lead to an optimistic assessment of the corrected field-scale performance. This design constrains our ability to quantify the sensitivity of forecasting skill to interannual hydroclimatic variability, and to evaluate performance across the full annual cycle, particularly during the non-growing season. Extending the analysis to longer, full-year records that span a broader range of climate states, together with independent calibration and evaluation of the field-scale correction, will be necessary for a rigorous assessment of model robustness.

4.3. Agricultural water management implications of multi-scale reference evapotranspiration forecasting

ET_o is a key atmospheric-demand input for crop-coefficient-based ET estimation (Fernández, 2023). Under the FAO56 framework, $ET_{c\ act}$ is commonly estimated by combining ET_o with appropriate $K_{c\ act}$, with additional adjustments used to represent water stress and other non-standard field conditions. In this study, these field conditions were represented using the empirically derived $K_{c\ act}$. Accurate ET_o forecasts can therefore provide useful lead-time information for agricultural water management when combined with locally representative $K_{c\ act}$ estimates.

The practical value of this forecasting framework depends on both forecast lead time and spatial scale, and should be judged relative to simple benchmarks. At the regional scale, the CNN-LSTM forecasts reproduced the seasonal evolution of ET_o but did not provide significant additional skill beyond climatology at any lead. For multi-week water-allocation planning, climatological ET_o therefore provides information of comparable or higher accuracy at minimal cost, and data-driven forecasts would add value only where they demonstrably outperform this benchmark.

At the vineyard scale, the agreement between forecasts and observations decreased more rapidly with lead time, indicating that field-scale application (e.g., $ET_{c\ act}$ estimation) requires site-specific correction and frequently updated $K_{c\ act}$ information. The ET_o^P showed closer agreement with the H-ET_o than with the PM-ET_o at both vineyards. This agreement should be interpreted cautiously because the Hargreaves equation does not explicitly represent wind speed or humidity and may be biased under windy conditions (Gavilán et al., 2006). Therefore, its application as a reference or input for ET_o forecasting at other sites should be supported by local validation and, where necessary, regional calibration.

The evaluated field-scale performance also depended on the observational reference. Whether direct EC observations or energy-closure-corrected estimates yielded better agreement differed between the two vineyards, and the choice of reference changed not only the magnitude but also the sign of the evaluated bias. This site dependence is consistent with the site-dependent closure behavior reported by Wang et al. (2024), even when advective effects were considered, and may be related to differences in water management. The field-scale representation of RIP720 combined two severely water-stressed and two deficit-irrigated treatments, whereas RIP760 was maintained under little to no water stress. Averaging the energy-balance components from four towers with contrasting water-stress levels may have combined treatment-specific differences in energy partitioning and closure, further complicating the comparison between the two references at RIP720. Together, these results indicate that uncertainty in EC-based ET, particularly that associated with energy-balance closure, can be comparable to forecast errors and should be considered when field-scale ET forecasts are evaluated or used for irrigation decisions; reporting results against both references, as done here, provides a more transparent assessment. Furthermore, when multiple towers represent contrasting management conditions within a forecast grid cell, future evaluations should compare the coarser-resolution forecasts with each tower individually rather than with their average, despite the scale mismatch. Such tower-level comparisons would preserve treatment-specific differences in water stress and energy partitioning, and could reveal how forecast performance varies with management conditions within the same grid cell, information that is lost when observations are averaged.

Nevertheless, the site dependence of $ET_{c\ act}$ forecast performance was based on only two vineyard sites and a single evaluation season, which limits the generalizability of this finding. Evaluation across additional sites, crop types (e.g., almond and pistachio), and growing seasons will be needed to determine whether such site-specific behavior persists under a wider range of environmental and management conditions. The appropriate forecast horizon should therefore be determined by the spatial scale, accuracy requirement, and decision context of the intended application, as well as by the uncertainty of the observations used for evaluation.

4.4. Actual crop coefficient as an additional constraint on coupled actual evapotranspiration forecasting

The empirical forecast of $K_{c\ act}$ further constrained the performance of coupled $ET_{c\ act}$ forecasts. From an operational perspective, $K_{c\ act}$ was derived from a remote-sensing-based $ET^O_{c\ act}$ and ET_o, rather than from $ET^{EC}_{c\ act}$ (or $ET^{EC,BR}_{c\ act}$), because the goal of this study was to develop a transferable framework that could be applied beyond instrumented research sites. However, both quantities used to derive $K_{c\ act}$ contain spatial representativeness uncertainty, which can propagate into its estimation and forecasting.

Specifically, the three datasets involved in the coupled framework, namely the CIMIS ET_o, the remotely sensed $ET^O_{c\ act}$, and EC flux tower estimates, represent different spatial supports. For the denominator of $K_{c\ act}$, the systematic differences between the 2-km CIMIS ET_o and the site-based PM-ET_o indicate that the gridded product did not fully represent the atmospheric demand at the vineyard sites. For the numerator, the differences between the $ET^O_{c\ act}$ and $ET^{EC}_{c\ act}$ (or $ET^{EC,BR}_{c\ act}$) indicate that $ET^O_{c\ act}$ may not fully represent the tower flux footprint. These mismatches can be incorporated into $K_{c\ act}$, increasing its variability and contributing to site-dependent forecast performance.

Future work should address four priorities to improve the accuracy and transferability of coupled $ET_{c\ act}$ forecasting.

The first priority is to quantify the uncertainty of the CIMIS ET_o product, which underlies the entire coupled framework. ET_o served not only as the input, training target, and evaluation reference for the CNN-LSTM, but also as the denominator of $K_{c\ act}$, so that its uncertainty propagates into every subsequent step. Beyond the comparison of the gridded ET_o with station-based PM-ET_o, assessing the uncertainty of the individual meteorological variables used to compute ET_o (e.g., air temperature and wind speed) would help identify which components contribute most to the ET_o uncertainty and under what conditions. Such an assessment would provide a basis for estimating the uncertainty of both the ET_o forecasts and the derived $K_{c\ act}$.

The second priority for future work is to reduce footprint mismatch among the datasets used to estimate and evaluate $K_{c\ act}$. Because the CIMIS ET_o, the finer-resolution $ET_{c\ act}$ product, and the dynamic EC flux footprint represent different spatial supports, their effective areas of

representation are difficult to align. These differences can obscure whether the variability in $K_{c\ act}$ reflects actual crop conditions or spatial mismatch among the input datasets. Developing higher-resolution ET_o products and footprint-aware spatial matching methods is therefore important for representing field-scale atmospheric demand and improving crop-coefficient estimation.

The third priority is to improve the estimation of $K_{c\ act}$ through more observation-driven remote-sensing methods. Deriving $K_{c\ act}$ from modeled $ET_{c\ act}$ products transfers uncertainties associated with model structure, forcing data, parameterization, and spatial scaling into the resulting $K_{c\ act}$. Evaluating and reducing these combined uncertainties can also require substantial effort from researchers. Remote-sensing approaches based more directly on observable crop properties may reduce dependence on modeled $ET_{c\ act}$. For example, multispectral imagery and point-cloud data combined with machine-learning methods can capture changes in crop growth stage, canopy development, and related structural characteristics more directly (Köksal et al., 2025; Singh Rawat et al., 2019). The uncertainty and transferability of these approaches across sites and crops still need to be evaluated.

The fourth priority is to improve the forecasting of $K_{c\ act}$ over future lead times. The LRU method used in this study relies on a single previous-year analog and was not compared with alternative forecasting approaches. Therefore, part of the limited forecast performance of $K_{c\ act}^{LRU}$ and $ET_{c\ act}^{LRU}$ may reflect limitations of the method itself rather than only the intrinsic predictability of $K_{c\ act}$. Unlike ET_o, which is governed mainly by meteorological conditions, $K_{c\ act}$ is strongly affected by field management, such as irrigation scheduling, imposed water stress, and canopy trimming, so that climatology- or persistence-based benchmarks may not adequately represent its expected behavior. Future research is therefore needed to develop empirical and machine-learning approaches that explicitly incorporate field management information and climatic drivers, and to establish appropriate benchmarks for evaluating $K_{c\ act}$ forecasts.

5. Conclusions

This study developed a CNN-LSTM framework for multi-week ET_o forecasting across regional and vineyard scales and evaluated it against climatology and anomaly persistence. A 60-day input sequence performed best, whereas longer inputs, including the CWT-derived 93-day timescale, did not improve performance, indicating the CWT-derived dominant timescale does not necessarily correspond to the timescale over which ET_o is predictable.

The CNN-LSTM reproduced the seasonal evolution of ET_o and corrected most of the bias of its raw persistence baseline, but it did not significantly outperform climatology at any lead, and its skill in forecasting ET_o anomalies was limited to the first two lead days. These results show that conventional accuracy metrics can overstate forecast usefulness for strongly seasonal variables and that benchmark-relative evaluation is essential.

Coupling forecasted ET_o with $K_{c\ act}$ extended the framework to crop water use, but $K_{c\ act}$ forecasting and the choice of EC reference introduced additional, site-dependent uncertainty. Future work should (1) quantify the uncertainty of the gridded ET_o product and its meteorological inputs, (2) reduce spatial mismatch among the ET_o, $ET_{c\ act}$, and EC datasets, (3) develop observation-driven remote-sensing methods for estimating $K_{c\ act}$, and (4) improve and benchmark $K_{c\ act}$ forecasting, supported by longer multi-year records.

CRediT authorship contribution statement

Rui Gao: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Mohammad Safeeq:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **John T. Abatzoglou:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **Joshua H. Viers:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization.

Declaration of Competing Interest

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

Acknowledgements

This work was supported in part by the US Department of Commerce, Economic Development Agency (EDA) Build Back Better Regional Challenge "Fresno-Merced Future of Food (F3) Initiative" award #07-79-07913 and the Agriculture and Food Research Initiative Competitive Grant No. 2021–69012–35916 from the USDA National Institute of Food and Agriculture to UC Merced's Secure Water Future (SWF) project. JTA and JHV were partially supported by the AI Research Institutes program jointly funded by NSF and USDA-NIFA through the AI Institute: Agricultural AI for Transforming Workforce and Decision Support (AgAID), award No. 2021–67021–35344. The authors also acknowledge technical support provided by the E&J Gallo scientific teams for data collection, organization, and analysis, as well as the cooperation of the vineyard management staff for logistical support and coordination of field operations with the GRAPEX team. The authors further acknowledge Alfonso Torres-Rua, William P. Kustas, Joseph G. Alfieri, Lynn McKee, John H. Prueger, Nicolas Bambach, Sebastian Castro, and Lawrence E. Hipps for their contributions to GRAPEX. Mention of trade names or commercial products is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the USDA. USDA is an equal opportunity provider and employer. We sincerely thank the Editor and anonymous reviewers for their thoughtful and constructive comments, which helped us substantially improve the focus, structure, and scientific rigor of this manuscript.

Appendix

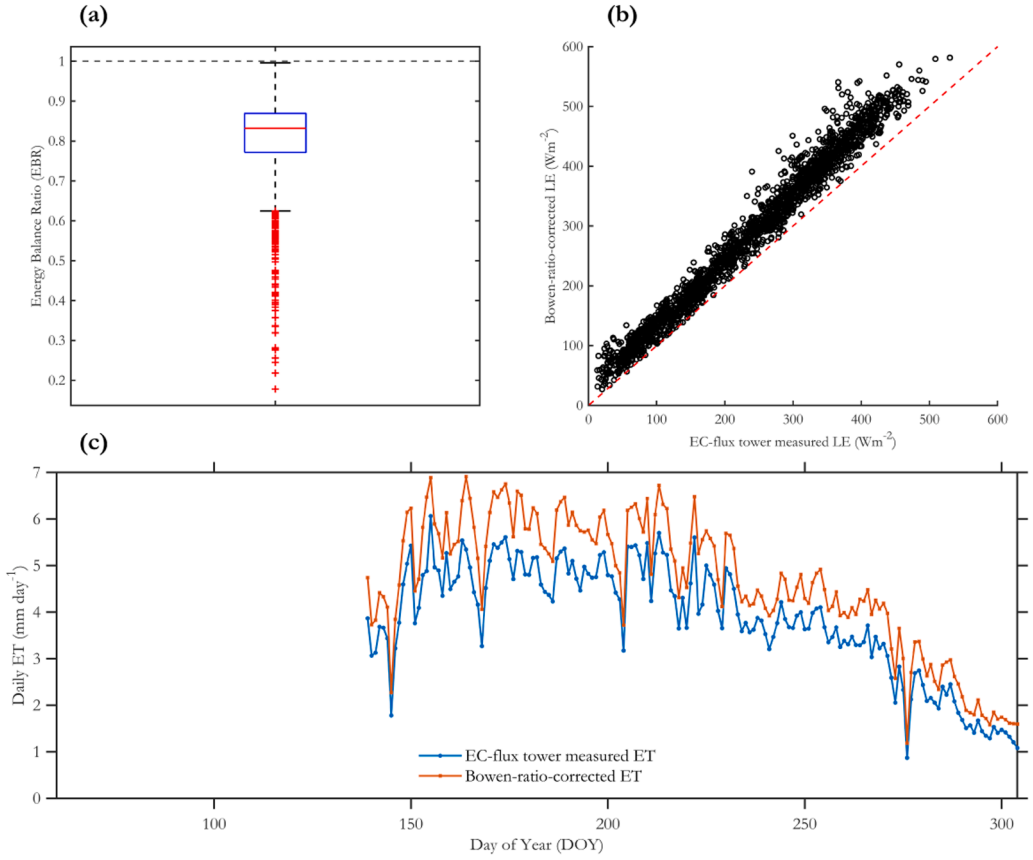

Fig A1. Energy balance ratio (EBR) distribution and EC-derived and Bowen-ratio-corrected latent heat flux (LE; W•m^{-2}) and daily ET (mm•day^{-1}) at the RIP720 vineyard. Panels (a) to (c) show the EBR distribution, hourly LE relationship, and daily ET time series, respectively. The horizontal dashed line in panel (a) indicates an EBR of 1, and the red dashed line in panel (b) indicates the 1:1 relationship

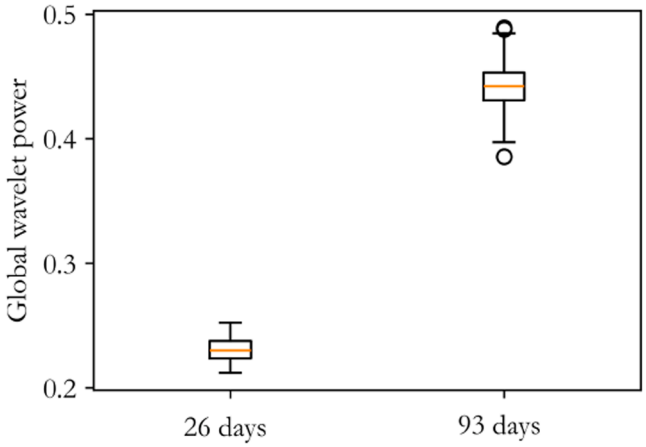

Fig A2. Distribution of global wavelet power at approximately 26- and 93-day scales across pixels within the fixed MOI

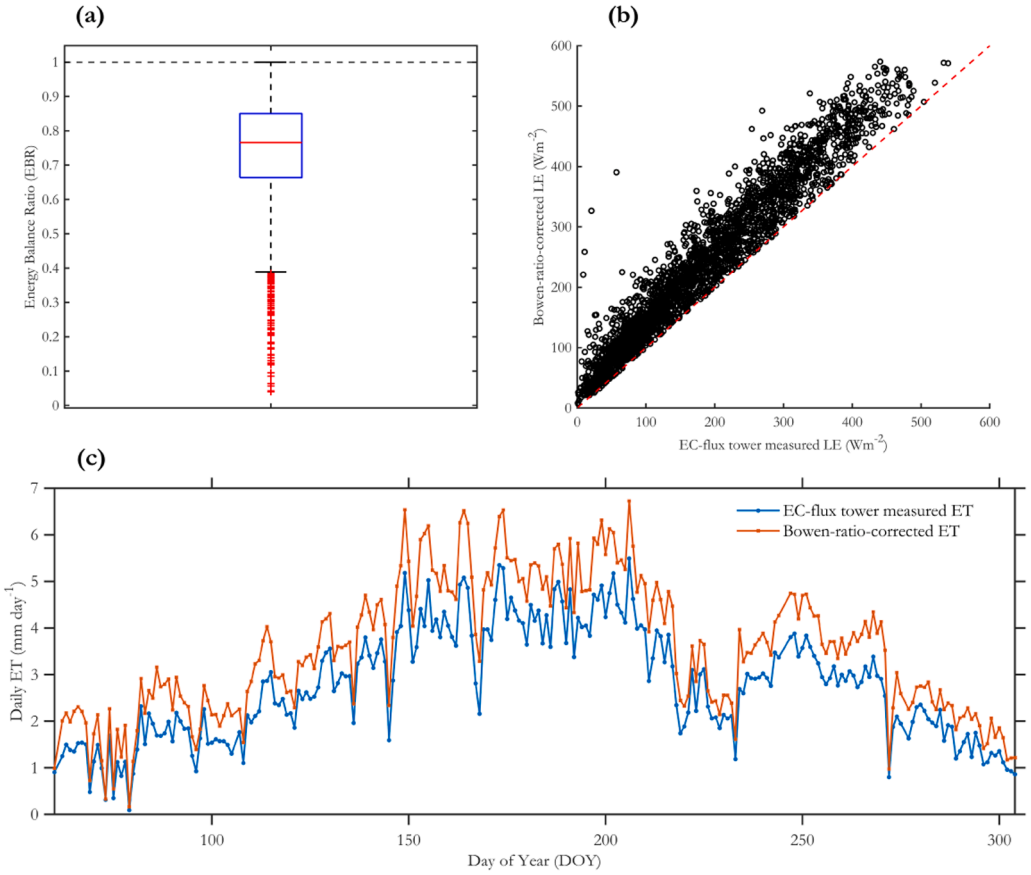

Fig A3. Energy balance ratio (EBR) distribution and EC-derived and Bowen-ratio-corrected latent heat flux (LE; W•m^{-2}) and daily ET (mm•day^{-1}) at the RIP760 vineyard. Panels (a) to (c) show the EBR distribution, hourly LE relationship, and daily ET time series, respectively. The horizontal dashed line in panel (a) indicates an EBR of 1, and the red dashed line in panel (b) indicates the 1:1 relationship

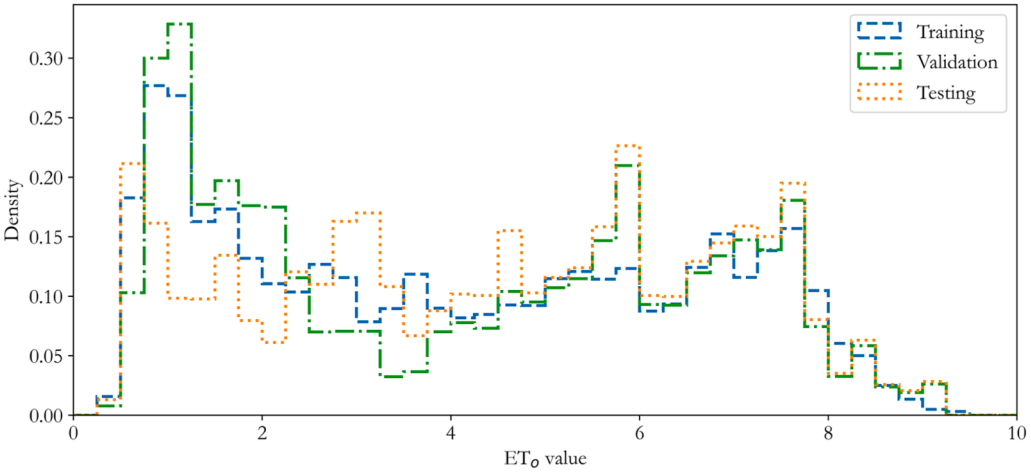

Fig A4. Distribution of daily ET_o in the training, validation, and testing sets, showing that the three splits have comparable ranges

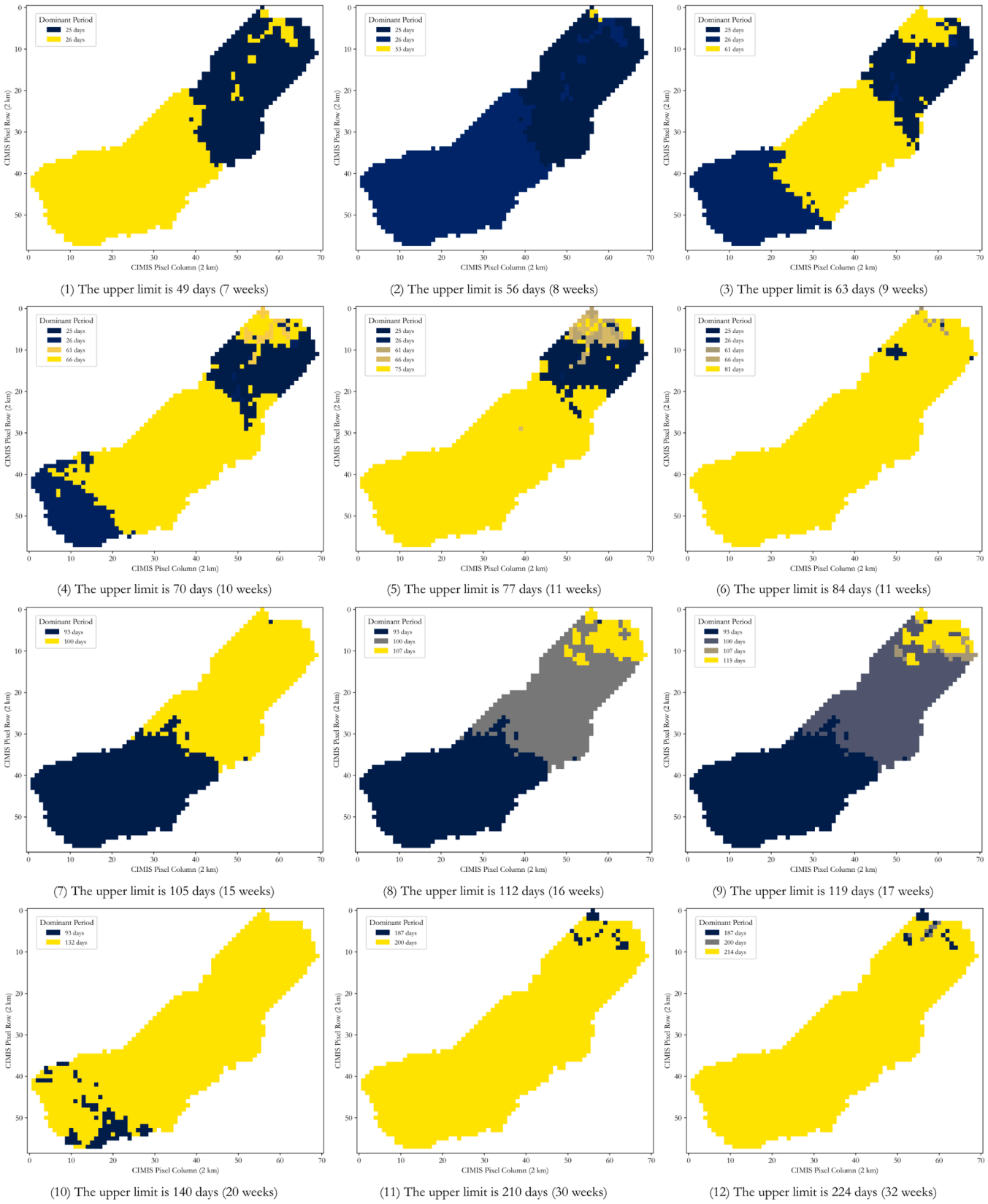

Fig A5. CWT results with different upper limits

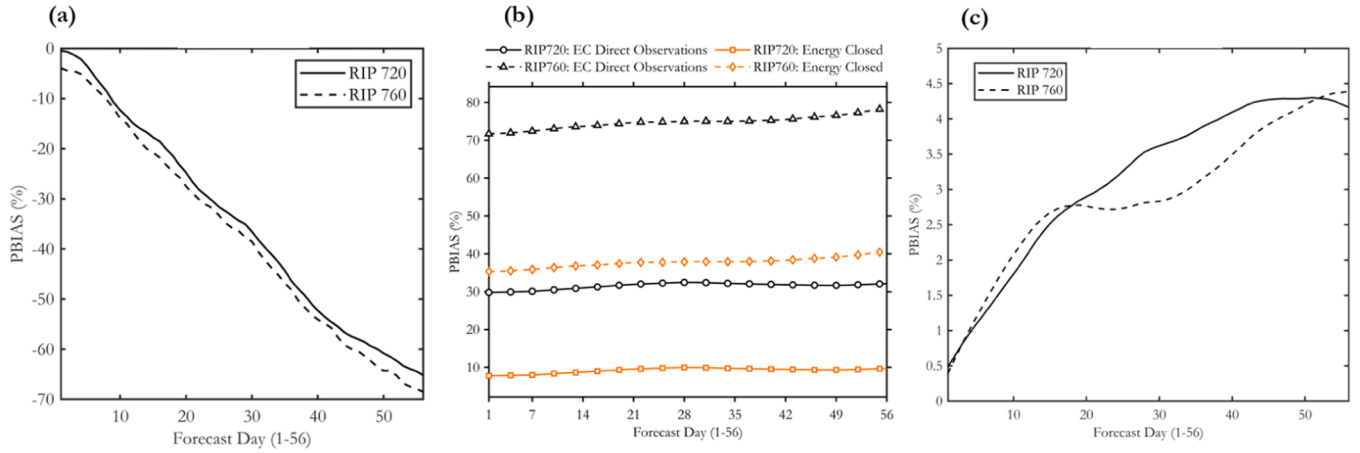

Fig A6. PBIAS across forecast lead days 1–56 for the forecasting analyses presented in Fig. 8 to Fig. 10. Panel (a) shows PBIAS for the locally corrected ET_o forecasts at RIP720 and RIP760 corresponding to Fig. 8. Panel (b) shows PBIAS for the $ET_{c\ act}$ forecasts corresponding to Fig. 9, evaluated against both $ET_{c\ act}^{EC}$ and $ET_{c\ act}^{EC,BR}$ at RIP720 and RIP760. Panel (c) shows PBIAS for the $K_{c\ act}^{LRU}$ forecasts at RIP720 and RIP760 corresponding to Fig. 10

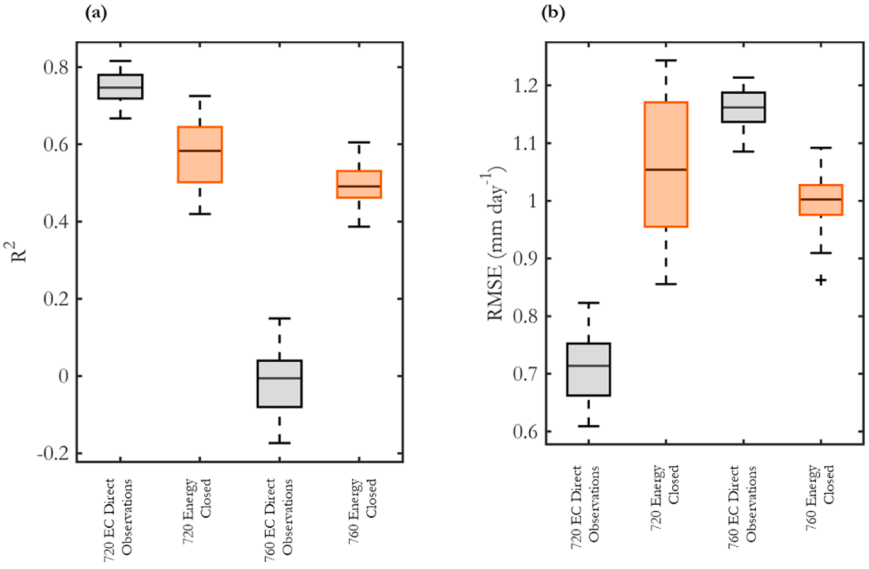

Fig A7. Boxplot summary of coupled $ET_{c\ act}$ forecast performance across the 56-day forecast horizon at RIP720 and RIP760. (a) Distribution of R^2 values. (b) Distribution of RMSE (mm•day^{-1}) values

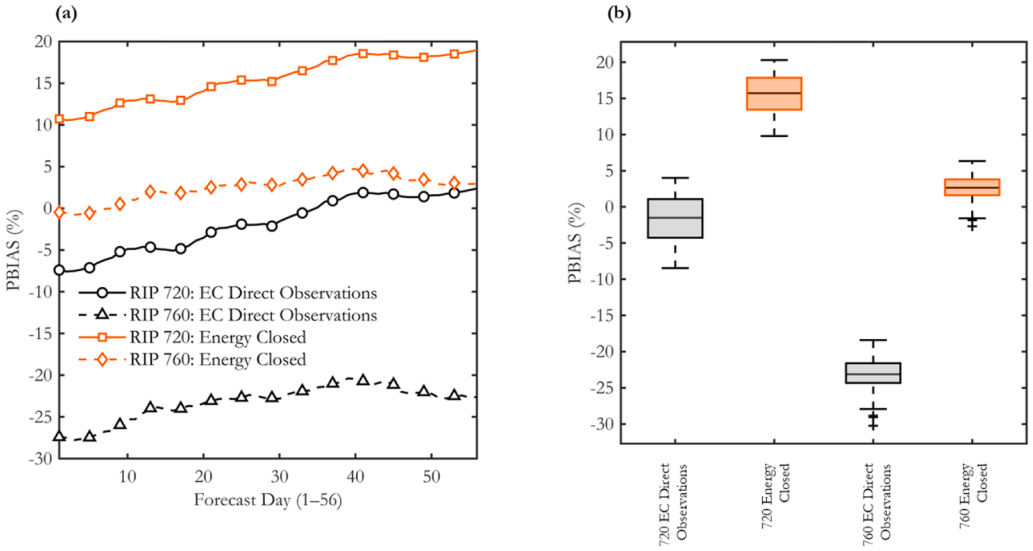

Fig A8. PBIAS of coupled $ET_{c\ act}$ forecasts evaluated against direct and energy-closure-corrected EC estimates at RIP720 and RIP760. Panel (a) shows changes in PBIAS across forecast lead days 1–56, and panel (b) summarizes the corresponding distribution across all 56 lead days for each site and observational reference

Data availability

Data will be made available on request.

References

Abatzoglou, J.T., McEvoy, D.J., Nauslar, N.J., Hegewisch, K.C., Huntington, J.L., 2023. Downscaled subseasonal fire danger forecast skill across the contiguous United States. Atmos. Sci. Lett. 24, e1165. https://doi.org/10.1002/asl.1165.

Abatzoglou, J.T., Williams, A.P., 2016. Impact of anthropogenic climate change on wildfire across western US forests. Proc. Natl. Acad. Sci. 113, 11770–11775. https://doi.org/10.1073/pnas.1607171113.

Adamowski, K., Prokoph, A., Adamowski, J., 2009. Development of a new method of wavelet aided trend detection and estimation. Hydrol. Process. 23, 2686–2696. https://doi.org/10.1002/hyp.7260.

Ahmadi, A., Daccache, A., Sadegh, M., Snyder, R.L., 2023. Statistical and deep learning models for reference evapotranspiration time series forecasting: A comparison of accuracy, complexity, and data efficiency. Comput. Electron. Agric. 215, 108424. https://doi.org/10.1016/j.compag.2023.108424.

Ahmadi, A., Kazemi, M.H., Daccache, A., Snyder, R.L., 2024. SolarET: A generalizable machine learning approach to estimate reference evapotranspiration from solar radiation. Agric. Water Manag. 295, 108779. https://doi.org/10.1016/j.agwat.2024.108779.

Albers, J.R., Newman, M., 2019. A priori identification of skillful extratropical subseasonal forecasts. Geophys. Res. Lett. 46 (21), 12527–12536. https://doi.org/10.1029/2019GL085270.

Allen, R.G., Walter, I.A., Elliott, R.L., Howell, T.A., Itenfisu, D., Jensen, M.E., Snyder, R. L., 2005. The ASCE Standardized Reference Evapotranspiration Equation. American Society of Civil Engineers, Reston, VA. https://doi.org/10.1061/9780784408056.

Altunkaynak, A., Ozger, M., 2016. Comparison of Discrete and Continuous Wavelet–Multilayer Perceptron Methods for Daily Precipitation Prediction. J. Hydrol. Eng. 21 (7), 04016014. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001376.

Anderson, M.C., Kustas, W.P., Norman, J.M., Diak, G.T., Hain, C.R., Gao, F., Yang, Yun, Knipper, K.R., Xue, J., Yang, Yang, Crow, W.T., Holmes, T.R.H., Nieto, H., Guzinski, R., Otkin, J.A., Mecikalski, J.R., Cammalleri, C., Torres-Rua, A.T., Zhan, X., Fang, L., Colaizzi, P.D., Agam, N., 2024. A brief history of the thermal IR-based Two-Source Energy Balance (TSEB) model – diagnosing evapotranspiration from plant to global scales. Agric. For. Meteorol. 350, 109951. https://doi.org/10.1016/j.agrformet.2024.109951.

Ben Hamouda, G., Zaccaria, D., Bali, K., Snyder, R.L., Ventura, F., 2022. Evaluation of Forecast Reference Evapotranspiration for Different Microclimate Regions in California to Enable Prospective Irrigation Scheduling. J. Irrig. Drain. Eng. 148, 04021061. https://doi.org/10.1061/(ASCE)IR.1943-4774.0001632.

Casirati, S., Conklin, M.H., Nandi, S., Safeeq, M., 2025. Effect of Forest Management Practices on Water Balance Across a Water–Energy Gradient in the Upper Kings River Basin, USA. Ecohydrology 18 (1), e2753. https://doi.org/10.1002/eco.2753.

Casirati, S., Conklin, M.H., Safeeq, M., 2026. Hydrological Response to Compounding Impacts of Climate Change and Forest Management in the Upper Kings River Basin, CA, USA. Ecohydrology 19 (1), e70157. https://doi.org/10.1002/eco.70157.

Fernández, E., 2023. Editorial note on terms for crop evapotranspiration, water use efficiency and water productivity. Agric. Water Manag. 289, 108548. https://doi.org/10.1016/j.agwat.2023.108548.

Gao, R., 2026b. met_gfit: A Python toolkit for comprehensive model evaluation and diagnostic plotting (Version v1.1.4) [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.20513771.

Gao, R., 2026a. A Python Toolkit for Reference Evapotranspiration Calculation Directly from Pandas DataFrames (Version Initial) [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.19197914.

Gao, R., Safeeq, M., Viers, J.H., 2026c. A lightweight Python package to download OpenET evapotranspiration (ET) time-series data (daily and monthly) using the OpenET API (Version 0.1.0) [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.18880781.

Gao, R., Torres-Rua, A.F., Nieto, H., Zahn, E., Hipps, L., Kustas, W.P., Alsina, M.M., Bambach, N., Castro, S.J., Prueger, J.H., Alfieri, J., McKee, L.G., White, W.A., Gao, F., McElrone, A.J., Anderson, M., Knipper, K., Coopmans, C., Gowing, I., Agam, N., Sanchez, L., Dokoozlian, N., 2023. ET Partitioning Assessment Using the TSEB Model and sUAS Information across California Central Valley Vineyards. Remote Sens. 15, 756. https://doi.org/10.3390/rs15030756.

Gao, R., Safeeq, M., Torres-Rua, A.F., Alfieri, J.G., Kustas, W.P., Prueger, J.H., Bambach-Ortiz, N., Castro, S., Hipps, L.E., McKee, L.G., Abatzoglou, J.T., Viers, J.H., 2025. Forecasting Evapotranspiration Products Using Convolutional Neural Network-Long Short-Term Memory (CNN-LSTM) to Support Agricultural Water Management in California Vineyards. ESS Open Archive (Preprint). https://doi.org/10.22541/essoar.176677235.59774814/v1.

Gao, R., Alsina, M.M., Torres-Rua, A.F., Hipps, L., Kustas, W.P., Anderson, M., Nieto, H., McElrone, A.J., Knipper, K., Bambach-Ortiz, N., Castro, S.J., Prueger, J.H., Alfieri, J., McKee, L.G., White, W.A., Gao, F., Coopmans, C., Gowing, I., Agam, N., Sanchez, L., Dokoozlian, N., 2026. A machine learning framework for California vineyard water status monitoring using sUAS Imagery and short-term meteorological data. Irrig. Sci. 44, 60. https://doi.org/10.1007/s00271-026-01102-8.

Gao, R., Safeeq, M., Viers, J.H., 2026d. A Python and Shell-based Toolkit for Spatial CIMIS Data Acquisition and Geospatial Preprocessing (Version initial) [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.19137639.

Gavilán, P., Lorite, I.J., Tornero, S., Berengena, J., 2006. Regional calibration of Hargreaves equation for estimating reference ET in a semiarid environment. Agric. Water Manag. 81, 257–281. https://doi.org/10.1016/j.agwat.2005.05.001.

Gómez, F., Lagos, O., Gil, P.M., Jara, J., Zaccaria, D., 2024. Assessment of reference evapotranspiration forecasting in the Mediterranean climate of Central Chile using the ASCE standardized Penman-Monteith equation, the Hargreaves-Samani equation, and weather predictions from the Global Forecast System model. Sci. Total Environ. 957, 177606. https://doi.org/10.1016/j.scitotenv.2024.177606.

Granata, F., Di Nunno, F., 2021. Forecasting evapotranspiration in different climates using ensembles of recurrent neural networks. Agric. Water Manag. 255, 107040. https://doi.org/10.1016/j.agwat.2021.107040.

Gupta, D., Hazarika, B.B., Berlin, M., 2020. Robust regularized extreme learning machine with asymmetric Huber loss function. Neural Comput. Appl. 32, 12971–12998. https://doi.org/10.1007/s00521-020-04741-w.

Hargreaves, G.H., Allen, R.G., 2003. History and Evaluation of Hargreaves Evapotranspiration Equation. J. Irrig. Drain. Eng. 129, 53–63. https://doi.org/10.1061/(ASCE)0733-9437(2003)129:1(53).

Hart, Q.J., Brugnach, M., Temesgen, B., Rueda, C., Ustin, S.L., Frame, K., 2009. Daily reference evapotranspiration for California using satellite imagery and weather station measurement interpolation. Civ. Eng. Environ. Syst. 26, 19–33. https://doi.org/10.1080/10286600802003500.

Jensen, M.E., Allen, R.G., 2016. Evaporation, Evapotranspiration, and Irrigation Water Requirements. American Society of Civil Engineers, Reston, VA. https://doi.org/10.1061/9780784414057.

Jha, P.K., Pathak, T.B., 2024. Seasonal climate forecasts show skill in predicting winter chill for specialty crops in California. Commun. Earth Environ. 5, 485. https://doi.org/10.1038/s43247-024-01623-0.

Kandhway, A., Scarpare, F., Liu, M., Nelson, R., Adam, J.C., Anderson, R.G., Conklin, M. H., Safeeq, M., 2025. Water use dynamics of almond and pistachio crops in the Mediterranean region amid climate change. Agric. Water Manag. 307, 109219. https://doi.org/10.1016/j.agwat.2024.109219.

Köksal, E.S., Tunca, E., Taner, S.C., 2025. Determination of Kc Curves of Wheat, Barley, Sunflower, Maize (Grain), Sugar Beet, and Onion by using Landsat 8–NDVI Time Series and Relationships between Kc and NDVI, in Bahadir, M., Haarstrick, A., Güneş, K., Topkaya, B., Al-Omari, A. (Eds.), Agricultural Water Management. Springer, Cham, pp. 85–102. https://doi.org/10.1007/978-3-031-96134-2_6.

Kukal, M.S., Allen, R.G., Kilic, A., Blankenau, P., DeJonge, K.C., Thorp, K.R., Huntington, J., Pereira, L.S., López-Urrea, R., Salman, M., Paredes, P., Robison, C., 2026. Artificial intelligence/machine learning estimation of FAO and ASCE standardized reference ET erodes the physical basis and intended purpose of ET standardization. Agric. Water Manag. 333, 110657. https://doi.org/10.1016/j.agwat.2026.110657.

Künsch, Hans, R., 1989. The jackknife and the bootstrap for general stationary observations. Ann. Stat. 17, 1217–1241. https://doi.org/10.1214/aos/1176347265.

Kustas, W.P., Agam, N., Ortega-Farias, S., 2019. Forward to the GRAPEX special issue. Irrig. Sci. 37, 221–226. https://doi.org/10.1007/s00271-019-00633-7.

Kustas, W.P., McElrone, A.J., Agam, N., Knipper, K., 2022. From vine to vineyard: the GRAPEX multi-scale remote sensing experiment for improving vineyard irrigation management. Irrig. Sci. 40, 435–444. https://doi.org/10.1007/s00271-022-00816-9.

Li, Z., Mu, Z., Gao, R., 2024. Applicability of ERA5 Reanalysis Precipitation Data in Runoff Modeling in China's Ili River Basin. J. Hydrol. Eng. 29 (5), 04024036. https://doi.org/10.1061/JHYEFF.HEENG-6161.

Liu, X., Hu, Z., Wang, Y., Wang, M., Hou, W., 2025. Spatial scale effects on the response strength of water resources supply-demand balance to driving factors. Ecol. Indic. 173, 113431. https://doi.org/10.1016/j.ecolind.2025.113431.

Murphy, A.H., 1988. Skill scores based on the mean square error and their relationships to the correlation coefficient. Mon. Weather Rev. 116. (12) 2417-2424. https://doi.org/10.1175/1520-0493(1988)116<2417:SSBOTM>2.0.CO;2.

Murphy, A.H., 1992. Climatology, Persistence, and Their Linear Combination as Standards of Reference in Skill Scores. Weather Forecast 7, 692–698. https://doi.org/10.1175/1520-0434(1992)007<0692:CPATLC>2.0.CO;2.

Murphy, A.H., Epstein, E.S., 1989. Skill scores and correlation coefficients in model verification. Mon. Weather Rev. 117 (3) 572-582. https://doi.org/10.1175/1520-0493(1989)117<0572:SSACCI>2.0.CO;2.

Nassar, A., Torres-Rua, A., Kustas, W., Alfieri, J., Hipps, L., Prueger, J., Nieto, H., Alsina, M.M., White, W., McKee, L., Coopmans, C., Sanchez, L., Dokoozlian, N., 2021. Assessing Daily Evapotranspiration Methodologies from One-Time-of-Day sUAS and EC Information in the GRAPEX Project. Remote Sens. 13, 2887. https://doi.org/10.3390/rs13152887.

Parker, D.D., Zilberman, D., 1996. The use of information services: The case of CIMIS. Agribusiness 12, 209–218. https://doi.org/10.1002/(SICI)1520-6297(199605/06)12:3<209::AID-AGR2>3.0.CO;2-4.

Pereira, L.S., Paredes, P., Hunsaker, D.J., López-Urrea, R., Mohammadi Shad, Z., 2021a. Standard single and basal crop coefficients for field crops. Updates and advances to

the FAO56 crop water requirements method. Agric. Water Manag. 243, 106466. https://doi.org/10.1016/j.agwat.2020.106466.
Pereira, L.S., Paredes, P., López-Urrea, R., Hunsaker, D.J., Mota, M., Mohammadi Shad, Z., 2021b. Standard single and basal crop coefficients for vegetable crops, an update of FAO56 crop water requirements approach. Agric. Water Manag. 243, 106196. https://doi.org/10.1016/j.agwat.2020.106196.
Pino-Vargas, E., Taya-Acosta, E., Ingol-Blanco, E., Torres-Rúa, A., 2022. Deep Machine Learning for Forecasting Daily Potential Evapotranspiration in Arid Regions, Case: Atacama Desert Header. Agriculture 12, 1971. https://doi.org/10.3390/agriculture12121971.
Safre, A.L.S., Nassar, A., Torres-Rua, A., Aboutalebi, M., Saad, J.C.C., Manzione, R.L., de Castro Teixeira, A.H., Prueger, J.H., McKee, L.G., Alfieri, J.G., Hipps, L.E., Nieto, H., White, W.A., del Mar Alsina, M., Sanchez, L., Kustas, W.P., Dokoozlian, N., Gao, F., Anderson, M.C., 2022. Performance of Sentinel-2 SAFER ET model for daily and seasonal estimation of grapevine water consumption. Irrig. Sci. 40, 635–654. https://doi.org/10.1007/s00271-022-00810-1.
Sharma, D., Sharma, A., Panda, S.K., Babel, M.S., Kumar, M., 2024. Wavelet analysis of rainfall and application of hydrological model in a semi-arid river basin of Rajasthan, India. Clean-Soil Air Water 53. https://doi.org/10.1002/clen.202300223.
Shukla, S., Safeeq, M., AghaKouchak, A., Guan, K., Funk, C., 2015. Temperature impacts on the water year 2014 drought in California. Geophys. Res. Lett. 42, 4384–4393. https://doi.org/10.1002/2015GL063666.
Singh Rawat, K., Kumar Singh, S., Bala, A., Szabó, S., 2019. Estimation of crop evapotranspiration through spatial distributed crop coefficient in a semi-arid environment. Agric. Water Manag. 213, 922–933. https://doi.org/10.1016/j.agwat.2018.12.002.
Sorooshian, S., Li, J., Hsu, K., Gao, X., 2011. How significant is the impact of irrigation on the local hydroclimate in California's Central Valley? Comparison of model results with ground and remote-sensing data. J. Geophys. Res. 116, D06102. https://doi.org/10.1029/2010JD014775.
Torres, A.F., Walker, W.R., McKee, M., 2011. Forecasting daily potential evapotranspiration using machine learning and limited climatic data. Agric. Water Manag. 98, 553–562. https://doi.org/10.1016/j.agwat.2010.10.012.
Trilnick, I., Huang, A., Silver, J., Gordon, B., Zilberman, D., 2022. The multiple benefits of CIMIS—publicly provided weather and irrigation information in California. ARE Update 25, 5–8. https://giannini.ucop.edu/filer/file/1661186505/20471/.
Wang, T., Alfieri, J., Mallick, K., Arias-Ortiz, A., Anderson, M., Fisher, J.B., Girotto, M., Szutu, D., Verfaillie, J., Baldocchi, D., 2024. How advection affects the surface energy balance and its closure at an irrigated alfalfa field. Agric. For. Meteorol. 357, 110196. https://doi.org/10.1016/j.agrformet.2024.110196.
Wilks, D.S., 2019. Statistical Methods in the Atmospheric Sciences, 4th ed. Elsevier, Amsterdam. https://doi.org/10.1016/C2017-0-03921-6.
Yang, R., Mu, Z., Gao, R., Huang, M., Zhao, S., 2024. Interactions between ecosystem services and their causal relationships with driving factors: a case study of the Tarim River Basin, China. Ecol. Indic. 169, 112810. https://doi.org/10.1016/j.ecolind.2024.112810.