Timescale-Guided Framework for Deep Learning Forecasting of CIMIS ET_o and Implications for Vineyard ET_a via K_a Coupling

Abstract: Accurate multi-week forecasting of evapotranspiration (ET) at different spatial scales can support water allocation to lower the risk management of extreme weather event (e.g., drought) for agricultural areas, yet reliable forecast actual ET (ET_a) remains challenging because it depends on both atmospheric demand and time-varying crop conditions. Most forecasting studies focus on reference ET (ET_o) at single locations or short lead times, and fewer evaluate how intrinsic ET_o timescales and input sequence design influence forecast horizons, or how uncertainty in crop coefficients limits ET_a forecasting. Here we develop and test a hybrid ET_a forecasting framework that couples (1) a CNN–LSTM model for multi-lead daily CIMIS ET_o forecasting with (2) empirical "environmental memory" approaches to forecast the actual crop coefficient (K_a), enabling ET_a estimation as $ET_a \approx K_a \times ET_o$. Using two vineyard sites in Madera County, California, we trained the CNN-LSTM on 2014 to 2017 CIMIS ET_o and evaluated forecasts using validation (2018-05-01 to 2018-07-31) and testing (2018-08-01 to 2018-10-31) periods. Wavelet analysis identified a dominant 93-day periodicity in ET_o during 2014-2017, providing a characteristic timescale for the region; however, model sensitivity experiments showed that this signal period does not directly correspond to the optimal input sequence length for deep learning. Comparing alternative historical windows (60 and 120 days) indicated improved performance with input sequences shorter than 93 days. Across both vineyard sites, ET_o forecast skill was stable through a 23-day horizon and remained acceptable at 30 days, demonstrating the feasibility of one-month ET_o forecasting. ET_a and K_a were less forecastable than ET_o, reflecting interannual variability in phenology, water status, and management. Ratio-based empirical K_a methods (Prevday-Ratio and Mean60-Ratio) outperformed a constant-coefficient formulation, while exhibiting systematic positive bias relative to modeled/derived ET_a and/or K_a. Overall, results highlight that ET_o timescale, input length, and forecast horizon should be co-designed, and that improved K_a dynamics – potentially supported by satellite vegetation indices – are critical for operational ET_a forecasting.

Keywords: Reference Evapotranspiration; Actual Evapotranspiration; Actual Crop Coefficient; Deep Learning; Forecasting

The abbreviations/symbols used in this study and their descriptions are listed below:

EC	Eddy covariance
CIMIS	California irrigation management information system
GRAPEX	Grape Remote sensing Atmospheric Profile and Evapotranspiration eXperiment
ET	Evapotranspiration
ET_o	Reference evapotranspiration
ET_a	Actual evapotranspiration
K_a	Actual/single crop coefficient
AOI	Area of interest
MOI	Modeling area of interest

CNN	Convolutional neural network
LSTM	Long short-term memory
CWT	Continuous wavelet transform
NDVI	Normalized difference vegetation index
MA	Moving average
R^2	Coefficient of determination
RMSE	Root mean square error
r	Pearson correlation coefficient
d	Willmott's index of agreement
L_{in}	Input window length

1. Introduction

Water scarcity and increasing hydroclimatic variability are intensifying the need for decision-support tools that enable proactive irrigation planning, particularly for perennial systems where operational choices often require weeks of lead time. This need is especially relevant in California's high-value perennial agriculture, where efficient irrigation planning under a changing climate requires understanding how phenology and biophysical processes translate into crop water use (Kandhway et al., 2025). Evapotranspiration (ET) is a key variable linking atmospheric demand to agricultural water use, yet forecasting ET at decision-relevant horizons remains challenging. From an operational perspective, around one month's lead time has been identified as a practical threshold for actionable forecasts (Shukla et al., 2015), and similar lead-time behavior has been reported for other specialty-crop climate services in California, where forecast skill is strongest at approximately one month and degrades at longer horizons (Jha and Pathak, 2024). Together, these findings motivate ET forecasting approaches that extend beyond short-range horizons and provide useful multi-week guidance for irrigation management.

A fundamental distinction in ET forecasting is between reference evapotranspiration (ET_o), which primarily reflects meteorological forcing, and actual evapotranspiration (ET_a), which additionally depends on crop condition, phenology, soil water status, and management. As a result, ET_a is typically harder to forecast robustly than ET_o. A common pathway to translate atmospheric demand into consumptive use is through coefficient coupling, where ET_a is expressed as a function of ET_o and the actual/single crop coefficient (K_a) expressed as $ET_a \approx K_a \times ET_o$ (Peddinti et al., 2025). However, the reliability of this approach depends on capturing time-varying K_a dynamics and ensuring scale-consistent inputs. In addition, gridded meteorological products (e.g., ET_o) can exhibit region- and season-dependent biases that propagate into downstream extimates, underscoring the need to evaluate forcing-data uncertainty when developing forecasting frameworks (Li et al., 2024). At the same time, obtaining high-spatial-resolution hydroclimatic products remains a persistent challenge for remote sensing applications, including drought monitoring (Gyaneshwar et al., 2023) and ET estimation, and spatial resolution limitations can propagate uncertainty into coefficient-based ET_a calculations.

In California, ET forecasting also links to broader water-security pressures. Sierra Nevada headwater watersheds are a vital source of water for ecosystems and the economy, yet warming-driven drought stress and wildfire risks are threatening landscape resilience and water-related services (Casirati et al., 2025). Recent modeling work shows that warming can alter ET differently across water- and energy-limited environments and intensify vegetation water stress, reinforcing the importance of anticipating ET dynamics under warmer conditions (Casirati et al., 2026). These hydroclimatic pressures increase the value of reliable, lead-time-aware ET information for agricultural water management, especially where planning horizons span multiple weeks.

Recent studies have mainly explored machine learning and deep learning approaches for ET_o forecasting, including open-source model evaluation frameworks across diverse climates. For example, Bellido-Jiménez et al. (2022) developed AgroML to benchmark machine learning and Transformer-based models for daily ET_o prediction and demonstrated the strong adaptability of data-driven models to agro-environmental time series. Nevertheless, much of the current literature focuses on station-based meteorological predictors and places limited emphasis on two issues that are central for operational adoption: (1) multi-week forecast horizons and (2) principled guidance on the amount of historical information required to support a given lead time. Several studies evaluate short forecasting windows (e.g., 7 days; Ben Hamouda et al., 2022; Gómez et al., 2024), but systematic assessment across longer horizons remains less common. More importantly, recent work has noted that practical guidance linking historical input length ("how much history") to achievable forecast horizon ("how far ahead") is largely missing in ET_o forecasting studies (Gao et al., 2025b). Addressing this gap is essential for designing parsimonious and robust forecasting models, especially when moving toward spatially distributed applications.

Wavelet analysis provides a physically interpretable approach to characterize dominant periodicities and multi-scale variability in hydrologic time series and has been widely used for feature extraction and variability assessment (Adamowski et al., 2009). Prior studies have demonstrated that wavelet-based representations can improve predictive modeling by isolating scale-dependent signals. For instance, Altunkaynak and Ozger (2016) coupled wavelet transforms with neural network models for precipitation forecasting, and Sharma et al. (2024) used continuous wavelet methods to diagnose evolving rainfall periodicities across seasons. These findings motivate the use of wavelet diagnostics not only as a preprocessing tool, but also as a mechanism to inform forecasting model design by identifying characteristic timescales that may govern predictability.

In this study, we develop a hybrid forecasting framework for irrigated perennial systems that combines (1) wavelet-based timescale diagnostics with (2) a CNN-LSTM model for multi-lead daily ET_o forecasting and (3) empirical "environmental memory" approaches for forecasting K_a, enabling ET_a estimation through coefficient coupling. Our objectives are to: (1) identify dominant temporal scales in ET_o and evaluate whether such scales align with effective historical input sequence lengths for forecasting, (2) quantify how ET_o predictive skill evolves across extended lead times to determine a decision-relevant forecast horizon, and (3) assess how alternative K_a forecasting formulations influence coupled ET_a performance while clarifying the implications of spatial resolution and representativeness in coefficient-based ETa estimation. By explicitly linking timescale, input length,

and forecast horizon, this work provides practical guidance for developing robust multi-week ET_o and ET_a forecasting models suitable for agricultural water management applications. This aligns with broader needs in California resource management, where timely decisions can be constrained by capacity and information availability; improving information support is recognized as a pathway to more effective management (Srinivasan et al., 2025).

2. Materials and Methods

2.1 Materials

This study was conducted in Madera County, California, located in the eastern San Joaquin Valley. The southwestern portion of the county is characterized by intensive agricultural production, where the ten most common crop types from the 2018 Land IQ dataset are shown (Figure 1). In contrast, the northeastern region is dominated by mountainous natural landscapes with relatively low human disturbance, and elevation gradually decreases from these uplands toward the agricultural lowlands; several lakes are also present near the agricultural region and contribute to the county's landscape. Two eddy-covariance (EC) flux tower sites established by the GRAPEX program (Kustas et al., 2022, 2019) were used as reference locations for model evaluation: RIP720, a "Merlot" vineyard, and RIP760, a "Chardonnay" vineyard. At RIP720, the vineyard included four management treatments described in Safre et al. (2022) and (Gao et al., 2025a, 2023), with one EC tower deployed for each treatment. For this study, we averaged the measured energy-balance components across the four towers and treated the aggregated fluxes as representative of the RIP720 site footprint. In addition to the county-scale Area of Interest (AOI), we defined a Modeling Area of Interest (MOI) as a rectangular domain (outlined by red dashed lines) that encompasses the primary agricultural region of Madera County (Figure 1). Details about the MOI are explained in Section 0.

We used three data sources in this study. First, the direct EC flux tower measurements collected at RIP720 and RIP760 in 2018 (dark green dash-dot lines in Figure 2a&c) were used as the primary reference for model evaluation. Second, the OpenET ensemble ET_a product for 2017-2018 at the EC-flux tower locations (gray solid lines in Figure 2a&c), which has been recommended for agricultural applications (Melton et al., 2022; Volk et al., 2024), was used in two ways: (1) to derive the K_a by pairing OpenET ET_a with CIMIS ET_o and then forecast the K_a using empirical methods, which was then combined with forecast ET_o to estimate future ET_a; and (2) to provide the sole input time series for the empirical ET_a forecasting methods. Third, daily gridded CIMIS ET_o from the California Irrigation Management Information System (CIMIS) was used both as the meteorological driver for ET_o forecasting and as an input for K_a derivation. We used the gridded CIMIS ET_o product over Madera County for 2014-01-01 to 2018-12-31 for the spatial ET_o forecasting experiments. For daily K_a calculations, we assumed that CIMIS ET_o is spatially uniform within the 2-km grid cell containing each EC flux tower. We therefore extracted the CIMIS ET_o time series at the grid cell containing each EC flux tower for 2017-01-01 to 2018-12-31 and paired it with the corresponding OpenET ensemble ET_a. The resulting K_a time series for each site is shown as the solid gray line in Figure 2b&d. The three empirical methods for ET_a and K_a forecasting are described in Section 2.2.2.

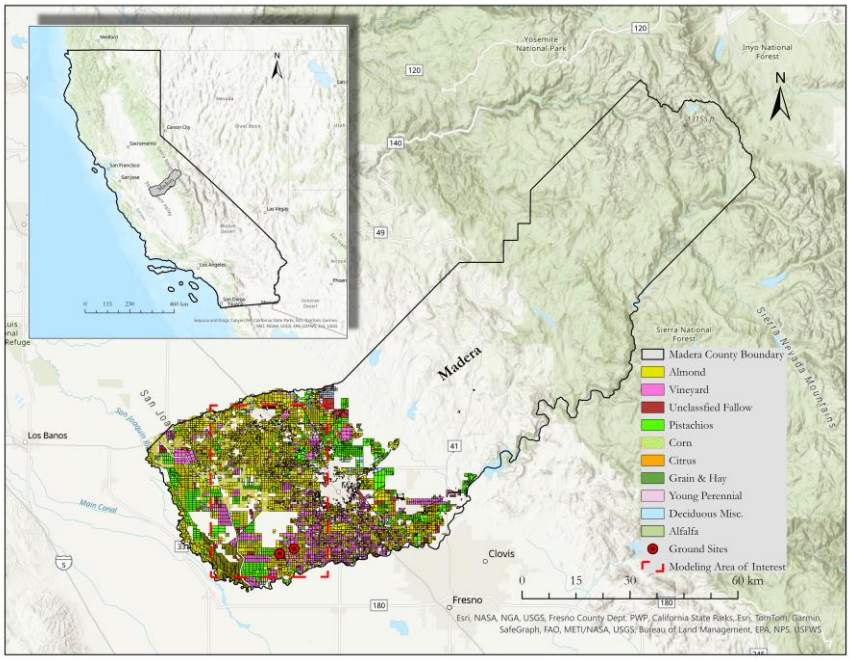


Figure 1.Study region and data sources in Madera County, California. The county boundary is shown on a hillshade basemap, with land-use/land-cover classes mapped for the agricultural area based on 2018 Land IQ data. The red dashed rectangle indicates the modeling area of interest (MOI) used for CNN-LSTM inputs (described in Section 0). The two eddy-covariance (EC) flux tower sites (RIP720 and RIP760) are marked and used for time-series extraction and model evaluation at the field scale. The inset map (upper left side) indicates the location of Madera County within California.

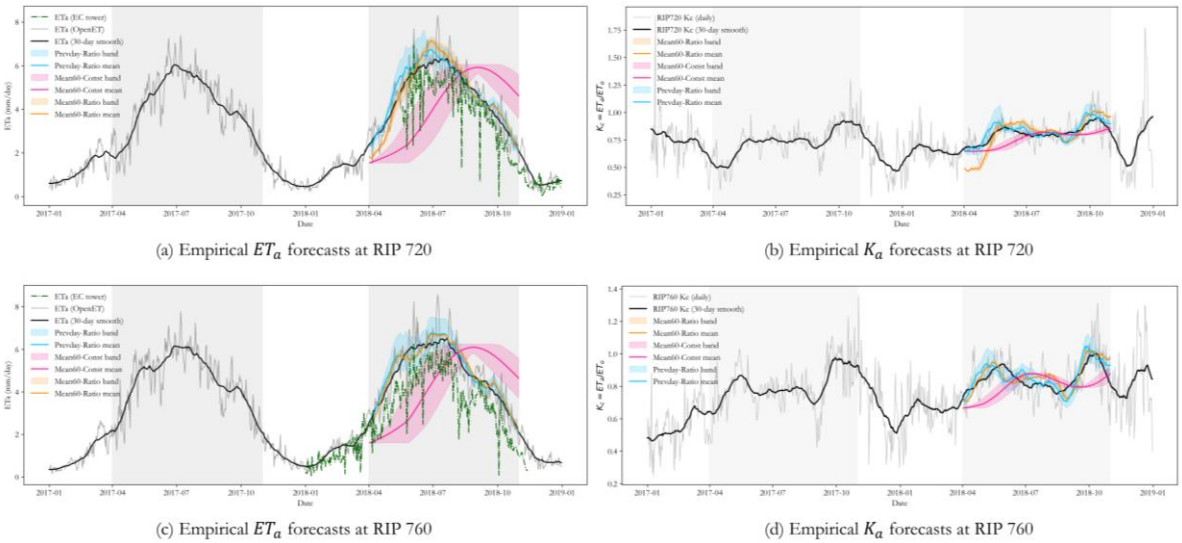


Figure 2. Empirical ET_a and K_a forecasts at two EC flux-tower vineyards. Panels (a) and (c) show daily OpenET ensemble ET_a (gray) and its 30-day smoothed series (black) at RIP720 and RIP760, respectively, together with EC tower ET_a observations (dark-green dash-dot). Colored curves and shaded envelopes represent the mean and prediction band of three empirical 56-day forecasting approaches (Mean60-Ratio, Mean60-Const, and Prevday-Ratio) during the 2018 growing season. Panels (b) and (d) show the corresponding K_a time series, including the daily and 30-day smoothed K_a from OpenET-CIMIS and the three empirical forecast methods. Light gray shading highlights the April-October periods in 2017 and 2018.

2.2 Methods

We implemented a workflow that links timescale diagnostics, ET_o forecasting design, and field-scale ET_a forecasting. First, we applied continuous wavelet transform (CWT) analysis to spatial daily CIMIS ET_o to identify dominant variability periods and to define candidate historical input window lengths (L_{in}) for forecasting. Second, we trained a CNN-LSTM model to forecast CIMIS ET_o at lead times of 1-56 days and used sensitivity experiments with input windows shorter and longer than the dominant period to test whether optimal model performance requires less history than the characteristic ET_o timescale. We then evaluated lead-day-specific ET_o skill across the full 56-day horizon to characterize how predictability decays with forecast lead time and to identify the practical range of reliable forecasts. Finally, we produced field-scale ET_a forecasts at two vineyard sites by coupling forecast ET_o with forecast crop response, where crop response was represented by the actual crop coefficient $K_a = ET_a^{OpenET}/ET_o^{CIMIS}$ and forecasted using three empirical "memory" methods; coupled ET_a forecasts were evaluated against EC tower ET_a.

2.2.1 CWT analysis of spatial CIMIS-ET_o and sensitivity analysis of the maximum period threshold

We adopted the CWT to decompose the gridded daily CIMIS ET_o between 2014-01-01 and 2017-12-31 into time-scale space and identify dominant sub-seasonal to seasonal cycles. Figure 3 shows the Morlet mother wavelet and its use in CWT. The CWT of a time series $x(t)$ is defined as Equation 1, where ψ is the mother wavelet, a is the scale parameter, and b is the translation parameter. The scale a controls the temporal width of the wavelet, enabling detection of either short-term or long-term patterns, with b shifts the wavelet along the time axis. The resulting coefficients $W(a,b)$ indicate how strongly the ET_o time series resembles the wavelet at each scale and time, capturing localized, scale-dependent features such as seasonal pulses or short-term fluctuations (Adamowski et al., 2009). In this study, we summarize the CWT coefficients using wavelet power $|W(a,b)|^2$, which represents local variance across scales and times.

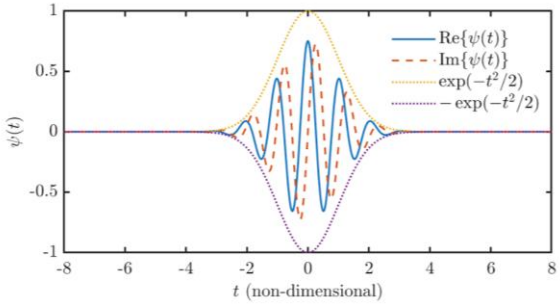


Figure 3. Illustration of the Morlet mother wavelet and its use in the continuous wavelet transform (CWT). The left panel shows the real and imaginary parts of the Morlet wavelet $x(t)$ as functions of non-dimensional time signal t. The oscillatory carrier is localized in time by the Gaussian envelope $e^{\frac{-t^2}{2}}$ and its negative counterpart, which bound the amplitude of the wavelet. The right panel gives the corresponding mathematical expressions: Equation (1) defines the CWT of a time series $x(t)$, Equation (2) defines the complex Morlet wavelet, and Equation (3) and (4) show its real and imaginary parts, respectively. In this study, this wavelet is used as the analyzing function to scan the daily ET_o time series across different scales and times.

$$W(a,b) = \frac{1}{\sqrt{a}} \int x(t)\psi(\frac{t-b}{a})dt$$ Equation 1

$$\psi(t) = \pi^{-1/4} \cdot e^{i\omega_0 t} \cdot e^{-t^2/2}$$ Equation 2

$$Re\{\psi(t)\} = \pi^{-1/4} \cdot \cos(\omega_0 t) \cdot e^{-t^2/2}$$ Equation 3

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

We used the complex Morlet wavelet, defined as Equation 2, which combines an oscillatory complex sinusoidal carrier $e^{i\omega_0 t}$ with a Gaussian envelope $\mathrm{e}^{-t^2/2}$ to ensure time localization. The central frequency ω_0 was set to 6, consistent with the default configuration in the $pycwt$ package. Because the Morlet wavelet is complex-valued, its real and imaginary components are from a cosine- and sine-modulated Gaussian pair (Equation 3 and Equation 4), allowing the CWT coefficients $W(a,b)$ to encode both amplitude and phase information. We used wavelet power to provide a single, interpretable measure for identifying the dominant timescale.

For each 2-km pixel, we applied the Morlet CWT to the daily CIMIS ET_o time series and identified dominant sub-seasonal to seasonal cycles. The CWT was computed using a complex Morlet mother wavelet over a target period band with a minimum period of 7 days and a maximum period up to 364 days (i.e., $P_{min} = 7$ days, $P_{max} \leq 364$ days). We used a daily sampling interval ($dt = 1$ day) and a scale resolution of $d_j = 0.1$, and summarized the transform using wavelet power $|W(a,b)|^2$. For each pixel, the global wavelet power spectrum was obtained by averaging power over time, and the dominant period was defined as the period corresponding to the maximum global power within the selected period band.

To assess how the estimated dominant period depends on the choice of P_{max}, we conducted a sensitivity analysis in which P_{max} was varied from 28 to 364 days (1-day increments), while keeping $P_{min} = 7$ days fixed. For each P_{max}, we recomputed the dominant period at every pixel by restricting the global power spectrum to the range 7~P_{max} days and select the period with maximum power. We then summarized each run by (1) counting the number of unique dominant periods across the domain (after rounding dominant periods to integer days for comparison) and (2) examining the spatial patterns of the resulting dominant-period maps in relation to the agricultural versus natural landscape gradients in Madera County.

2.2.2 Empirical forecasting of ET_a and K_a

To complement the deep-learning ET_o forecasts and to provide transparent baselines, we implemented three empirical forecasting methods to forecast (1) daily ET_a derived only from ensembled ET_a (Figure 2 a and c), and (2) K_a which is derived from CIMIS ET_o and OpenET ensemble ET_a (Figure 2 b and d) at the EC-flux tower location.

We derived daily K_a at each EC tower site by pairing OpenET ensemble ET_a with CIMIS ET_o at the corresponding 2-km pixel (Equation 5). Because the ratio can be noisy, we also computed a smoothed series using a moving-average window (default 30 days). Using the historical year (2017), we built a lead-day "ratio table" that maps each anchor day t to a set of multiplicative ratios $r_h(t)$ for each lead day h. The forecast K_a in 2018 is then obtained by applying these ratios to a baseline K_a available on the anchor day.

$$K_a(t) = \frac{ET_a(t)}{ET_o(t)} \qquad \text{Equation 5}$$

We evaluated three empirical approaches for forecasting ET_a; the same formulations were also used to forecast K_a by replacing ET_a with K_a in the equations below.

Mean60-Ratio

For each anchor day t, we compute a baseline ET_a as the mean of the previous 60 days (Equation 6), and define lead-day ratios using 2017 data (Equation 7). In 2018, ET_a forecasts are generated as (Equation 8). We adopted a 60-day window as a shorter-history reference relative to the ~93-day dominant ET_o timescale identified in Section 3.2.

$$\overline{ET_a}(t) = \frac{1}{60}\sum_{k=1}^{60} ET_a(t-k) \qquad \text{Equation 6}$$

$$r_h(t) = \frac{ET_a(t+h)}{\overline{ET_a}(t)} \qquad \text{Equation 7}$$

$$\widehat{ET_a}(t+h) = \overline{ET_a}(t) \cdot r_h(t) \qquad \text{Equation 8}$$

Prevday-Ratio

This method uses the most recent available ET_a value, $ET_a(t)$, at the anchor day as baseline, and constructs lead-day ratios in 2017 as (Equation 9). Then, the forecasts for 2018 are (Equation 10). Because $ET_a(t)$ can be replaced by a smoothed ET_a when desired, this approach can be interpreted as a "real-time updating" forecast driven by the most recent crop condition.

$$r_h(t) = \frac{ET_a(t+h)}{ET_a(t)} \qquad \text{Equation 9}$$

$$\widehat{ET_a}(t+h) = ET_a(t) \cdot r_h(t) \qquad \text{Equation 10}$$

Mean60-Const

Finally, we include the Mean60-Const method as the simplest baseline. It assumes no temporal evolution beyond recent conditions by setting ET_a for all lead days equal to its 60-day trailing mean at the anchor day t, without applying lead-dependent ratios (Equation 11). This represents a "no-change" forecast relative to the recent mean state.

$$\widehat{ET}_a(t+h) = \overline{ET}_a(t), h = 1, \dots, H. \quad \text{Equation 11}$$

2.2.3 Convolutional long short-term memory (CNN-LSTM)

After identifying the dominant time scales, we restricted CNN-LSTM development to the MOI (Figure 1). Specifically, we used a 19 × 13 grid at 2-km resolution covering the primary agricultural region, rather than the full 59 × 71 county-wide domain. This choice was motivated by two practical considerations: (1) the two ground observation sites used for evaluation are located within the MOI; and (2) the smaller, regular rectangular boundary improves computational efficiency for CNN-LSTM training and simplifies boundary handling for the convolutional neural network (CNN) component, especially because pixels outside the Madera County boundary were masked as NaN.

The gridded CIMIS ET_o (2014-01-01 to 2018-12-31) was split chronologically for model development and independent evaluation. Data from 2014-01-01 to 2017-12-31 were used for training, 2018-05-01 to 2018-07-31 for validation, and 2018-08-01 to 2018-10-31 for testing. Because the CNN-LSTM requires a L_{in}, the effective number of available sequences within a fixed calendar window decreases as L_{in} increases. To enable consistent comparison across experiments with different L_{in}, we kept the validation and testing calendar windows fixed and evaluated all models using the same periods (May – July 2018 for validation and August – October 2018 for testing), while allowing the number of valid sequences to vary accordingly.

A detailed description of the CNN-LSTM architecture, training procedure, and hyperparameter selection is provided in (Gao et al., 2025b; Sarwar et al., 2025). Here, we implemented a spatiotemporal CNN-LSTM model that forecasts daily ET_o over the MOI using a persistence-plus-residual design. For each input day, a single 2-D convolutional layer (3 × 3 kernel; followed by 2 × 2 max pooling) encodes the ET_o field into a compact spatial feature representation, which is flattened and passed to a single-layer LSTM to capture temporal dependencies across the input sequence. The last LSTM hidden state is decoded into a spatial field by a fully connected layer reshaped into a feature map, followed by a 1 × 1 convolution that produces multi-step forecasts (up to 56 lead days).

To stabilize long-horizon predictions, the model predicts bounded residuals (scaled tanh) added to a persistence baseline (repeating the last observed ET_o map across all lead days), constraining forecast deviations while preserving spatial variability. To improve robustness to extreme ET_o values, we explored loss reweighting strategies that upweight errors above fixed thresholds (e.g., 6 mm/day) and above high quantiles (p95/p99) using smooth sigmoid gates, as well as a weighted Huber loss to reduce sensitivity to outliers. Because these alternatives did not improve validation RMSE or lead-day skill relative to the standard mean squared error (MSE) in our setting, the final model was trained with the

Adam optimizer (with L2 weight decay) to minimize MSE, with gradient clipping applied to improve numerical stability.

We also 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, we filtered candidate configurations using an estimated input-to-parameter ratio computed from the training sample count, sequence length, the number of valid pixels in the MOI, and the total number of model parameters; configurations below a minimum ratio threshold were excluded from further evaluation. Forecast skill was evaluated for lead days 1-56 using a one-day stride.

2.2.4 Model evaluation metrics

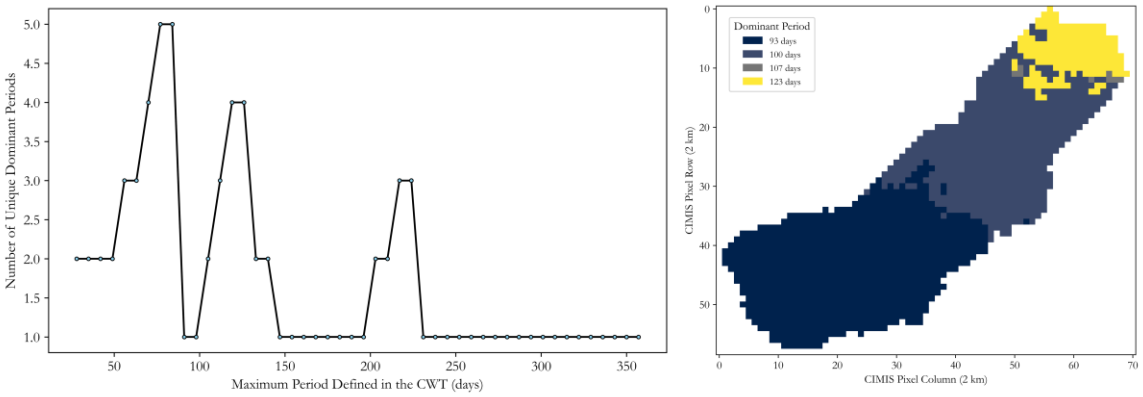
Model performance was evaluated using four standard statistics: coefficient of determination (R^2), root mean square error (RMSE), Pearson correlation coefficient (r), and Willmott's index of agreement (d). These metrics were computed for all experiments to quantify explained variance, error magnitude, linear association, and overall agreement between predictions/forecasts and observations. Implementation details (including optional diagnostic plots such as 1:1 scatter, residual, and Q-Q plots) are provided in the accompanying open-source Python repository archived on Zenodo (Gao, 2025).

3. Results

3.1 CWT analysis

Figure 4 (a) shows the sensitivity of the dominant-period mapping to the choice of the maximum period threshold (P_{max}), summarized as the number of unique dominant periods identified across the domain. Among the tested thresholds, P_{max}=126 days (18 weeks) was selected because it produced spatial patterns most consistent with land cover and management contrasts in Madera County. Representative dominant-period maps for 12 additional thresholds are provided in the Appendix (Figure A 1) for comparison.

Figure 4 (b) illustrates the spatial distribution of dominant ET_o cycles under this selected 126-day configuration. Four dominant periodicities (93, 100, 107, and 123 days) were identified, capturing both sub-seasonal and seasonal variability in daily ET_o. Spatially, shorter cycles (93 days) were concentrated in the southern and southwestern agricultural regions at lower elevations, whereas longer cycles (up to 123 days) were observed in the northern and northeastern regions at higher elevations. This variation closely corresponds to land-use patterns: the southern region, with intensively managed croplands, exhibited shorter, irrigation-influenced cycles, whereas the northern region, with less-managed areas, showed longer, climate-driven cycles.



(a) Sensitivity analysis to confirm the upper limit of the time window for the CWT method

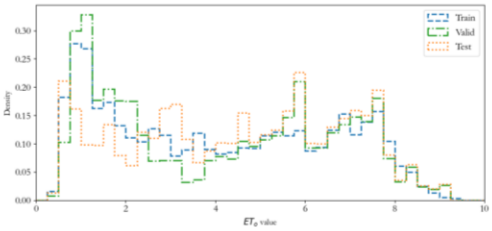
(b) One situation when the upper limit is 126 days (18 weeks)

Figure 4. (a) Sensitivity analysis of the CWT-based dominant-period mapping showing how the number of unique dominant periods across the domain changes as the maximum period threshold (P_{max}) increases from 28 to 364 days (7-day step). (b) Dominant-period map under the selected configuration (P_{max} = 126 days), highlighting four dominant periods (93, 100, 107, and 123 days). Shorter periods are concentrated in the southern agricultural region, whereas longer periods are more common in the northern natural areas.

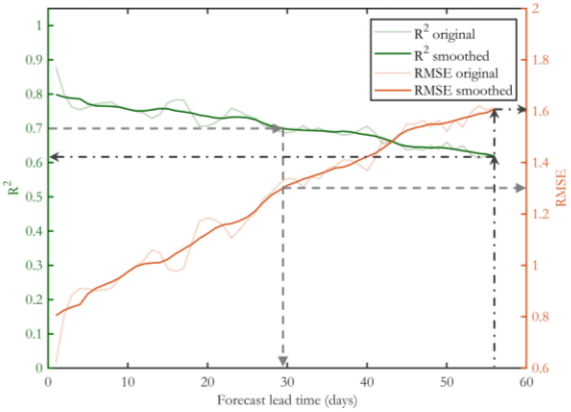
3.2 CNN-LSTM model performance within the area at different input window lengths

The train/validation/test splits were designed to be comparable in magnitude and variability of ET_o. as shown by the overlapping histograms in Figure 5 (a), the three subsets span similar ET_o ranges and exhibit broadly consistent distributions, which reduces the likelihood that differences in model skill are driven primarily by distribution shift across splits rather than by model structure or L_{in}.

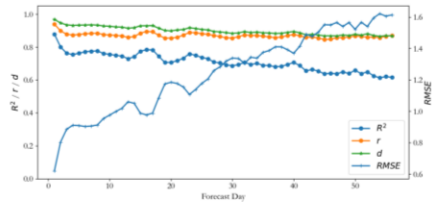
The lead-day evaluation curves in Figure 5 (c, d, and e) show a consistent degradation in forecast skill as lead time increases: RMSE increases with lead, while correlation (r), Willmott's d, and R^2 gradually decline. This behavior is expected because uncertainty accumulates with horizon and the predictability of daily atmospheric demand decreases from synoptic to subseasonal time scales. However, the three panels also reveal that L_{in} strongly affects how quickly skill decays. Although the CWT analysis identified a dominant ET_o periodicity of 93 days at the county scale, this result does not directly prescribe the optimal L_{in} for the CNN-LSTM. We hypothesized that an input window shorter than 93 days could be sufficient (and potentially advantageous) for multi-week forecasting. Consistent with this expectation, the model trained with L_{in}= 60 days achieved comparable or better performance across most lead days than models using L_{in} = 93 and L_{in} = 120. In particular, the L_{in} = 60 configuration maintained higher R^2 and lower RMSE at medium-to-long lead times, indicating that extending the input history beyond ~60 days did not yield additional predictive benefit for this task.



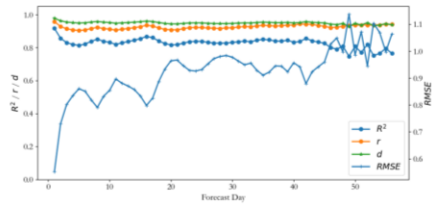
(a) Distribution of daily ET_o in the training, validation, and testing sets, showing that the three splits have comparable ranges.



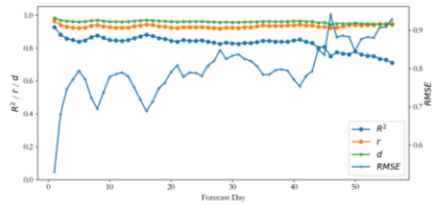
(b) Summary view for the L_{in} = 60 case showing R^2 (left y-axis) and RMSE (right y-axis) with lead day from 1 to 56 days. Faded curves report metrics computed from raw daily series, whereas vivid curves report metrics after applying a 7-day moving average. The gray dashed line marks R^2 = 0.70, corresponding to a lead day of ≈ 30 and RMSE of ≈ 1.30 mm/day. The gray dash–dot line marks lead day 56, where R^2 ≈ 0.62 and RMSE ≈ 1.61 mm/day.



(c) Lead-day evaluation metrics for CNN-LSTM models trained with input window length L_{in} = 60



(d) Lead-day evaluation metrics for CNN-LSTM models trained with input window length L_{in} = 93



(e) Lead-day evaluation metrics for CNN-LSTM models trained with input window length L_{in} = 120.

Figure 5. CNN-LSTM ET_o forecasting performance with different forecast lead time (1~56 days) and L_{in} (60, 93, and 120 days).

A closer look at the L_{in}= 60 summary (Figure 5, b) provides more interpretable benchmarks for long-horizon performance. When metrics are computed on the raw daily series, R^2 decreases steadily with lead day while RMSE increases. Applying a 7-day moving average to both forecasts and observations yield higher R^2 and lower RMSE across lead days, indicating that the model captures sub-weekly to weekly variability more reliably than day-to-day noise at long horizons. Using the smoothed curves as a practical skill indicator, the model reaches approximately R^2≈ 0.70 at a lead time of ~30 days with RMSE≈ 1.30 mm day^{-1}, and retains moderate skill at 56 days (R^2≈ 0.62, RMSE≈ 1.61 mm day^{-1}). These results suggest a useful forecasting regime up to roughly one month for higher-confidence applications, with diminishing but still informative skill extending to eight weeks.

Overall, Figure 5 indicates that the CWT-derived dominant period is valuable for constraining the plausible range of temporal dependencies, but it does not necessarily define the optimal L_{in} for supervised forecasting. Compared with L_{in}= 93 and 120 days, the L_{in}= 60 days delivered better ET_o forecasts in our experiments.

3.3 Evaluation of field-scale ET_a forecasting at EC-flux tower sites

3.3.1 Empirical ET_a forecasting skills

Empirical-method performance during the 2018 growing season (May-October) differed substantially among approaches (Figure 6). Across both RIP720 and RIP760, the Prevday-Ratio method most consistently reproduced observed ETa variability, indicating that anchoring forecasts to the most recent state better preserves short-term dynamics than methods based on longer-term averaging. This advantage is reflected in the lead-day evaluation (Figure 6a&b), where Prevday-Ratio generally produced higher R^2, r, and d, and lower RMSE than the other empirical baselines throughout the 56-day horizon. Mean60-Ratio tracked Prevday-Ratio closely but with slightly reduced skill, while Mean60-Const performed poorly: R^2rapidly became negative and RMSE increased with lead day, consistent with the limitation of assuming near-constant ET_a over an active growing season when irrigation and phenological development can drive rapid changes.

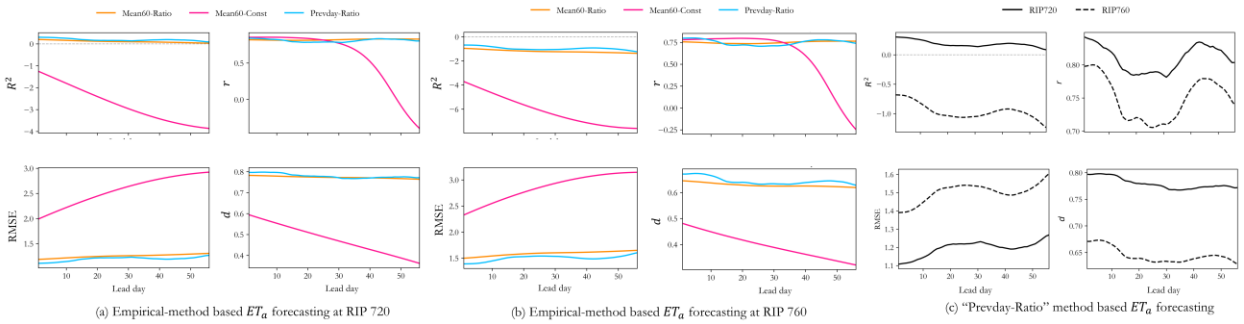


Figure 6. Lead-day evaluation of 56-day ET_a forecasting at two EC flux tower sites (RIP720 and RIP760). Panels (a) and (b) compare three empirical methods (Mean60-Ratio, Mean60-Const, and Prevday-Ratio) at each site. Panel (c) shows Prevday-Ratio performance for both sites. Forecast accuracy is quantified using R^2, RMSE (mm/d), Pearson correlation (r), and Willmott index of agreement (d).

Prevday-Ratio performance also differed between sites (Figure 6c). RIP720 maintained higher agreement with EC observations than RIP760 across most lead times, with a modest decline over the first ~two weeks followed by a slower rate of degradation through approximately one month. Beyond ~30 days, skill metrics exhibited greater variability (localized increases followed by declines), suggesting reduced robustness at longer horizons. This behavior may reflect intermittent management-driven shifts (e.g., irrigation events) and evolving canopy conditions that are not explicitly represented in the empirical memory formulation. Overall, these results indicate that empirical ET_a forecasting based solely on prior-year ratios is most reliable at short to mid-range lead times and becomes increasingly uncertain as the forecast horizon extends.

3.3.2 Empirical K_a forecasting skills

Empirical K_a forecasting skill varied markedly among methods during the 2018 growing season (Figure 7). At both RIP720 and RIP760, Prevday-Ratio shows the strongest overall agreement with the K_a series derived from OpenET ET_a and CIMIS ET_o, with generally higher R^2, r, and d than the two Mean60-based baselines across most lead times (Figure 7a&b). Mean60-Ratio exhibits a similar but

more stable (flatter) lead-time trajectory, with slightly reduced skill and reduced sensitivity to lead-day variability. In contrast, Mean60-Const deteriorates rapidly, with R^2 becoming negative at short lead times, indicating that a constant-coefficient assumption cannot represent the time-varying K_a dynamics during the active growing season.

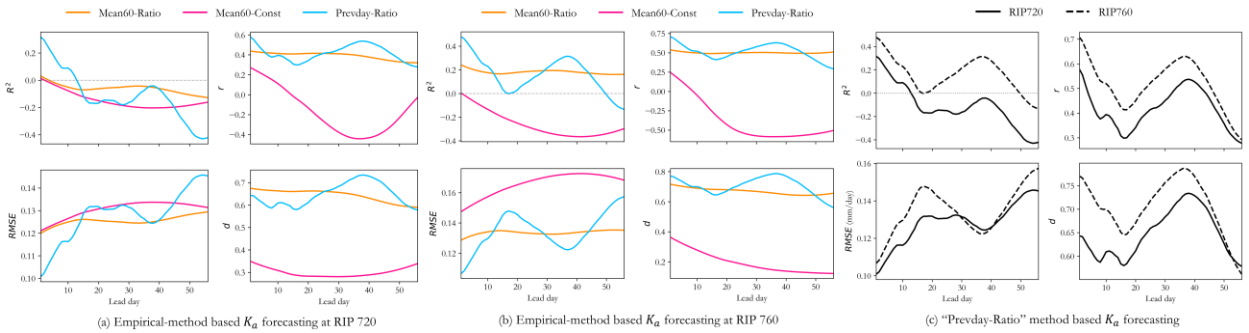
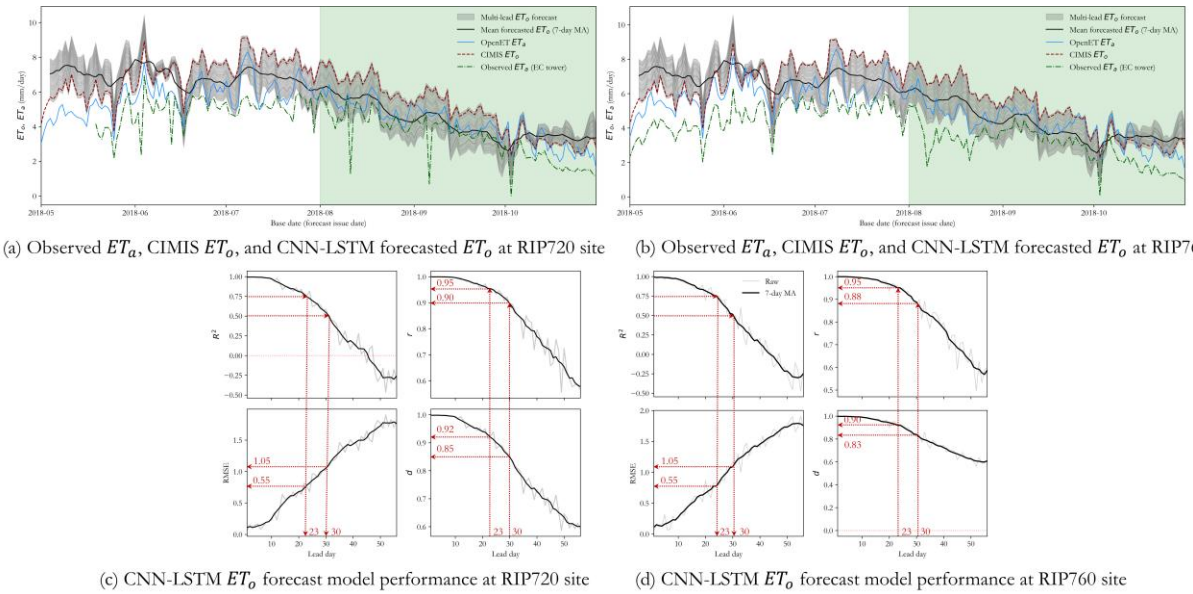


Figure 7. Lead-day evaluation of 56-day K_a forecasting at two EC flux tower sites (RIP720 and RIP760). Panels (a) and (b) compare three empirical methods (Mean60-Ratio, Mean60-Const, and Prevday-Ratio) at each site. Panel (c) shows Prevday-Ratio performance for both sites. Forecast accuracy is quantified using R^2, RMSE (dimensionless), Pearson correlation (r), and Willmott index of agreement (d).

Prevday-Ratio skill also shows non-monotonic evolution with forecast lead time (Figure 7c). After an initial decline over approximately the first 10 to 15 days, performance metrics exhibit a localized recovery around lead days ~35 to 40, accompanied by a temporary reduction in RMSE, followed by renewed degradation toward longer horizons. Notably, short-lead performance remains only moderate in terms of explained variance (R^2), even though correlation and agreement (r and d) remain relatively high, highlighting the difficulty of forecasting K_a magnitude and variability using prior-year ratio memory alone. Overall, ratio-based empirical approaches clearly outperform the constant baseline, but K_a predictability remains lead-time dependent and can introduce uncertainty into coupled ET_a forecasting as horizons extend.

3.3.3 CNN-LSTM ET_o forecasts EC-flux tower sites

The CNN-LSTM model achieved consistently high skill for daily CIMIS ET_o at both vineyard sites (Figure 8), indicating that atmospheric evaporative demand is comparatively predictable at multi-week lead times. Using the 7-day moving-average evaluation (black curves in Figure 8c&d), forecast performance remained strong through approximately 23 lead days, with $R^2 > 0.75$, RMSE< 0.55 mm/d, $r > 0.95$, and $d > 0.90$ at both RIP720 and RIP760. Skill declined with increasing horizon but remained meaningful at ~30 days, with R^2, r, and d near 0.50, 0.90, and 0.84, respectively, and RMSE around 1.05 mm/d. Overall, these results show that the CNN-LSTM provides a stable ET_o forecast component through short- to medium-range horizons, supporting its use as the atmospheric-demand driver in the coupled ET_a forecasting framework.



(a) Observed ET_a, CIMIS ET_o, and CNN-LSTM forecasted ET_o at RIP720 site

(b) Observed ET_a, CIMIS ET_o, and CNN-LSTM forecasted ET_o at RIP760 site

(c) CNN-LSTM ET_o forecast model performance at RIP720 site

(d) CNN-LSTM ET_o forecast model performance at RIP760 site

Figure 8. CNN-LSTM multi-lead forecasting performance for daily CIMIS ET_o at two vineyard sites. (a-b) Time-series comparison at RIP720 and RIP760. The gray envelope summarizes the distribution of multi-lead CNN-LSTM ET_o forecasts issued on each base date (forecast issue date), and the thick black line shows the mean forecast after applying a 7-day moving average to emphasize the synoptic-scale trend. CIMIS ET_o (red dashed) is the reference target for evaluation; OpenET ensemble ET_a (blue) and EC-tower ET_a (green dash–dot) are overlaid for context. Light-green shading denotes the independent test period (Aug-Oct 2018). (c–d) Lead-day skill of CNN-LSTM ET_o forecasts (lead days 1-56) at RIP720 and RIP760, quantified using R^2, RMSE (mm/d), Pearson correlation (r), and Willmott's index of agreement (d). Gray curves show metrics computed on raw daily series, and black curves show metrics computed after applying a 7-day moving average. Red dashed guides indicate representative lead times (e.g., ~23 and ~30 days) used to summarize short- versus medium-range forecast skill.

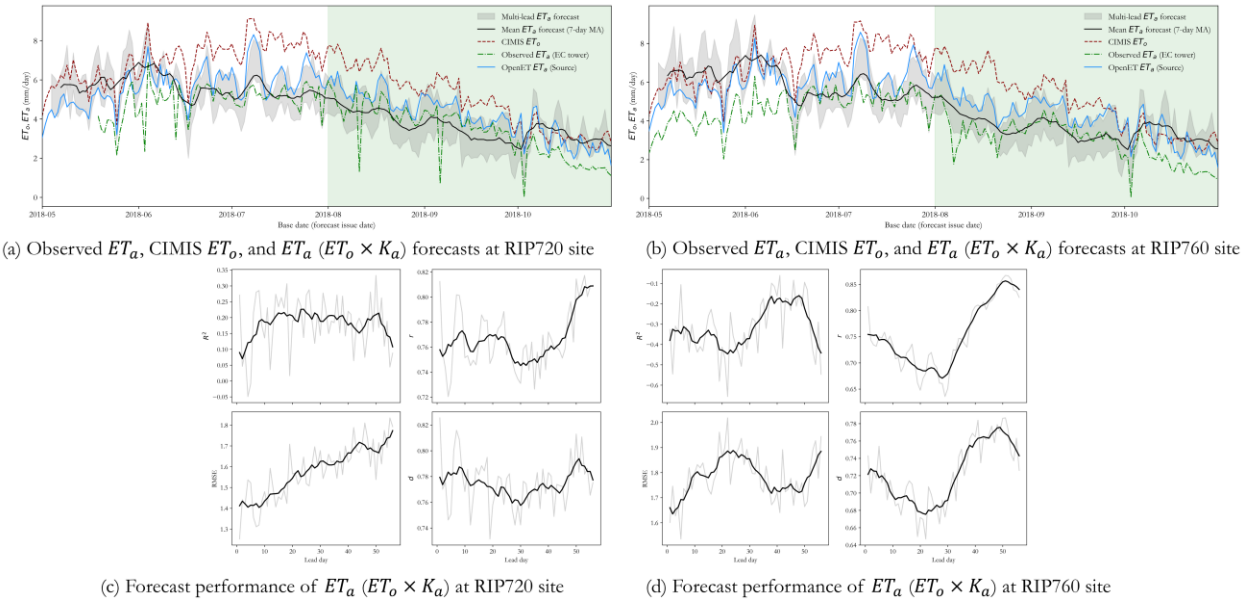
3.3.4 Field scale ET_a forecasting using CNN-LSTM ET_o and Prevday-Ratio K_a

Coupling CNN-LSTM ET_o forecasts with the Prevday-Ratio K_a formulation provides multi-lead field-scale ET_a forecasts at RIP720 and RIP760. Figure 9 summarizes coupled ET_a performance evaluated against independent EC flux-tower observations.

The coupled framework captures the overall seasonal decline in ET_a from August into early October (Figure 9a&b). For most of the growing season, the forecasted ET_a envelope (gray shading) remains below CIMIS ET_o, as expected for vineyards where ET_a is typically lower than reference demand. A systematic overestimation emerges in October, when coupled ET_a remains elevated relative to EC observations. This bias likely reflects the combined effects of (1) positive bias in the forecast ET_o component (Figure 8) and (2) overestimation in the empirically forecast K_a component (Figure 2b&d), together with a seasonal regime shift in which ET_a becomes increasingly constrained by phenology and canopy condition rather than atmospheric demand alone – processes not explicitly represented in the current empirical coupling.

Across lead times, RIP720 shows higher overall agreement than RIP760 (Figure 9c&d), indicating more reliable tracking of ET_a evolution at RIP720. However, daily-scale skill remains limited: R^2 remains below ~0.35 and RMSE exceeds ~1.3 mm/d, suggesting that a substantial fraction of day-to-day variance and magnitude is not captured. In contrast, Pearson r remains relatively high

(generally >0.72), indicating that the coupled forecasts reproduce the timing of rises and declines in ET_a even when they underperform in amplitude (bias and variability). This is also reflected by non-monotonic, lead-dependent behavior in r and Willmott's d.



(a) Observed ET_a, CIMIS ET_o, and ET_a ($ET_o \times K_a$) forecasts at RIP720 site

(b) Observed ET_a, CIMIS ET_o, and ET_a ($ET_o \times K_a$) forecasts at RIP760 site

(c) Forecast performance of ET_a ($ET_o \times K_a$) at RIP720 site

(d) Forecast performance of ET_a ($ET_o \times K_a$) at RIP760 site

Figure 9. Evaluation of field-scale ET_a forecasting via CNN-LSTM ET_o and Prevday-Ratio K_a coupling. (a-b) Time-series comparison at RIP720 and RIP760. The gray envelope summarizes the distribution of multi-lead ET_a forecasts obtained by coupling CNN-LSTM ET_o with site-specific empirical K_a. The thick black line shows the mean forecast after applying a 7-day moving average. CIMIS ET_o (red/magenta dashed; verify color), OpenET ensemble ET_a (blue), and EC flux-tower ET_a (green dash-dot) are overlaid for context. Light-green shading denotes the independent test period (Aug-Oct 2018). (c–d) Lead-day skill of coupled ET_a forecasts (lead days 1-56) at RIP720 and RIP760, quantified using R^2, RMSE (mm/d), Pearson correlation (r), and Willmott's index of agreement (d). Gray curves denote metrics computed on raw daily series, and black curves denote metrics after applying a 7-day moving average.

4. Discussions

The dominant 93-day periodicity identified for the southeastern Madera site is consistent with Sorooshian et al. (2011), who reported that irrigation-induced impacts on local weather and climate attenuate rapidly beyond irrigated areas. In our analysis, this 93-day signal represents a characteristic variability timescale in CIMIS ET_o for the region during 2014-2017. This spatially varying timescale is consistent with broader findings that environmental interactions often exhibit strong spatial structure and driver dependence (Yang et al., 2024). Importantly, however, a dominant signal period should not be conflated with an optimal L_{in} for CNN-LSTM training, because the sequence length that maximizes forecast skill also depends on model capacity, noise characteristics, and how much past information remains predictive of future dynamics. To evaluate whether the optimal L_{in} is likely shorter or longer than 93 days, we tested two additional candidate windows (60 and 120 days) as reference points. These comparisons suggest that model performance favors L_{in} shorter than 93 days, although identifying the optimal length will require a more systematic sensitivity analysis across a finer range of sequence lengths.

CNN-LSTM forecasting performance at the two vineyard sites (Figure 8) further indicates that predictive skill remains relatively stable through a 23-day forecast horizon and remains acceptable at 30 days. This result suggests that, under the conditions evaluated here, a 60-day input sequence may be sufficient to support approximately one-month ET_o forecasting. Taken together, the findings point to a potential coupling among (1) the characteristic ET_o timescale, (2) historical L_{in}, and (3) achievable forecast horizon, implying that these factors should be co-designed rather than selected independently. Future work should test this hypothesis explicitly using controlled experiments that vary L_{in} and forecast horizon jointly, while also expanding evaluation across additional sites and years to assess robustness and generalizability.

A key limitation of the present study is the relatively short temporal span used for model development and evaluation. We selected the 2014-2018 CIMIS ET_o record based on site observations and partitioned it into training (2014-2017), validation (2018-05-01 to 2018-07-31), and testing (2018-08-01 to 2018-10-31). Consequently, the model was trained on only four growing seasons, and testing was limited to a portion of the 2018 growing season (August-October). This design constrains our ability to quantify the sensitivity of forecasting skill to interannual hydroclimatic variability (including drought) 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 will be necessary for a rigorous assessment of model robustness.

Because ET_o is primarily governed by meteorological forcing, it is generally more forecastable than ET_a or K_a; similar conclusions have been reported in point-based ET_o forecasting at daily (Pino-Vargas et al., 2022) and monthly scales (Ahmadi et al., 2023). In contrast, ET_a and K_a reflect not only atmospheric demand but also time-varying canopy condition, phenology, water stress, and management, all of which can change substantially between seasons. This distinction is also consistent with the multiplicative structure of the coupled framework, in which $ET_a \approx K_a \times ET_o$: uncertainty and bias in either component – particularly in K_a – propagate directly into ET_a and can limit overall forecast skill even when ET_o forecasts are strong. Within this context, the Prevday-Ratio empirical approach shows promise for forecasting both ET_a and K_a in perennial systems under the assumption that commercial perennial crops are managed similarly across seasons. Nevertheless, Figure 2 shows that empirical forecasts tend to exceed modeled/derived ET_a and/or K_a, indicating a systematic positive bias. This bias plausibly reflects year-to-year shifts in phenology and irrigation-induced water status that alter K_a trajectories even within the same vineyard, and therefore limits predictability based solely on information from the preceding year.

An additional source of uncertainty arises from spatial-scale and representativeness mismatches in the K_a formulation, $K_a = ET_a^{OpenET}/ET_o^{CIMIS}$. In this study, the relatively coarse spatial resolution of CIMIS ET_o required us to assume spatially uniform reference demand within each ~2-km grid cell. By contrast, OpenET ET_a is resolved at the pixel/field scale and can vary substantially within the same area due to spatial heterogeneity in vegetation, soil moisture, and surface–atmosphere coupling. Because the numerator and denominator may not represent the same effective footprint, sub-grid variability in ET_o (e.g., driven by advection, local wind and humidity gradients, and terrain) can be

aliased into K_a, thereby inflating variance and contributing to spatially dependent bias in K_a predictability. These scale effects highlight the importance of obtaining higher-resolution ET_o (or developing robust downscaling approaches) to ensure consistency with field-scale ET_a products and to improve the reliability of K_a-based ET_a forecasting. Future work should quantify this uncertainty by testing alternative spatial sampling strategies (e.g., field-boundary averaging versus single-pixel extraction), evaluating sensitivity to ET_o interpolation/station representativeness, and comparing K_a derived from independent ET_a datasets when available.

We observed clear differences among empirical formulations even when they used the same historical window length. Mean60-Ratio closely follows Prevday-Ratio while exhibiting a narrower spread, indicating that averaging over a 60-day window can reduce day-to-day noise and stabilize ratio-based "memory" without fully eliminating responsiveness to changing conditions. In contrast, Mean60-Const performs poorly, consistent with the limitation of a constant-coefficient assumption that cannot represent time-varying crop and atmospheric controls on K_a. Together, these results suggest that ratio-based updating is essential for empirical K_a forecasting, and they motivate broader improvements through (1) integrating time-varying vegetation information (e.g., NDVI and LAI) to better track phenology and stress considering canopy condition strongly regulates surface energy partitioning and ET components (Gao et al., 2022c, 2022b, 2022a) and (2) expanding benchmarking across longer and more diverse open datasets to quantify transferability across years, sites, and management regimes.

At the same time, the choice of a 60-day window introduces an inherent responsiveness trade-off. Figure 2 shows that a trailing 60-day mean can induce substantial temporal lag, particularly for Mean60-Const, delaying K_a and ET_a adjustments during rapid transitions (e.g., early-season canopy development or management-driven shifts). A similar lag signature appears in Mean60-Ratio at RIP720 in early April, when predicted K_a and ET_a remain markedly below the corresponding smoothed baselines, implying that the trailing window is slow to capture the onset of seasonal change. These patterns indicate that a two-month memory may be overly long for K_a and ET_a dynamics at these sites. A practical next step is to treat the window length as a tunable parameter and identify an optimal range through systematic testing (e.g., grid search with cross-year validation), potentially allowing the window to vary seasonally. More advanced extensions could further reduce lag by constraining K_a evolution with remote-sensing indicators or learned K_a dynamics, although such approaches increase model complexity relative to the empirical baselines evaluated here.

In this study, K_a was derived from the OpenET ensemble ET_a and CIMIS ET_o rather than from EC-based ET_a because our goal is to develop a transferable framework that can be applied beyond instrumented research sites where flux towers are unavailable. This choice makes K_a derivation consistent with widely available operational datasets, enabling broader deployment across fields and regions. However, K_a computed from OpenET ET_a may differ from K_a computed using EC ET_a due to differences in spatial support, retrieval assumptions, and potential bias between remotely sensed and tower-based ET_a; these representativeness differences may also contribute to the site-dependent skill patterns observed in Figure 6c and Figure 7c. Accordingly, the K_a time series used here should be interpreted as an "OpenET-consistent" coefficient that is suitable for operational coupling but not necessarily equivalent to a tower-calibrated crop coefficient. A useful next step is to quantify these

differences directly by comparing OpenET-based and EC-based K_a at the tower sites and evaluating how the choice of K_a definition influences K_a forecasting skill and coupled ET_a performance.

Overall, our findings suggest that future near-term research should treat (at minimum) the intrinsic ET_o timescale, the selected L_{in}, and the forecast horizon as jointly interacting design variables, while simultaneously strengthening K_a estimation by integrating dynamic vegetation information and evaluating performance across full-year, multi-year records that include both drought and non-drought conditions.

5. Conclusion

This study developed and tested a hybrid multi-week forecasting framework for vineyard water use that combines CNN-LSTM prediction of CIMIS ET_o with empirical forecasting of the K_a to estimate field-scale ET_a. Wavelet analysis identified a dominant ~93-day timescale in CIMIS ET_o for southeastern Madera County (2014-2017); however, sensitivity experiments showed that this characteristic period does not correspond to the optimal historical L_{in} for deep learning. Instead, shorter L_{in} improved ET_o forecast performance, with skill remaining stable through ~23 lead days and still acceptable at ~30 days, supporting the feasibility of one-month ET_o forecasting for irrigation planning under the evaluated conditions.

In contrast, ET_a and K_a were less predictable than ET_o, reflecting the influence of phenology, water stress, and management on crop response. Ratio-based empirical approaches (Prevday-Ratio and Mean60-Ratio) outperformed a constant-coefficient baseline, but K_a estimates exhibited positive bias and limited transferability across years, which constrained ET_a forecasting skill. These results emphasize that reliable ET_a forecasting for agricultural water management will require improving K_a dynamics in addition to forecasting atmospheric demand.

Overall, the findings suggest that ET_o timescale, historical L_{in}, and forecast horizon should be treated as coupled design choices when developing multi-week forecasting models. Future work should (1) optimize empirical memory length through systematic sensitivity testing, (2) incorporate time-varying vegetation information (e.g., satellite indices) to better constrain K_a evolution, and (3) quantify scale and representativeness differences associated with K_a derived from operational ET products. Robustness should also be assessed using longer, full-year multi-year records spanning drought and non-drought conditions and additional sites to evaluate generalizability for operational deployment.

Acknowledgements

This work was supported in part by 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, and by

the Economic Development Administration Build Back Better Regional Challenge Farms-Food-Future Innovation Initiative (EDA #77907913). We also acknowledge exceptional technical support from the E&J Gallo scientific teams for data collection, organization, and analysis, and the cooperation of the vineyard management staff for logistical support and coordination of field operations with the GRAPEX team. 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 thank L. Parker for coordinating team science.

Reference

Adamowski, K., Prokoph, A., Adamowski, J., 2009. Development of a new method of wavelet aided trend detection and estimation, in: Hydrological Processes. pp. 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. https://doi.org/10.1016/j.compag.2023.108424

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

Bellido-Jiménez, J.A., Estévez, J., Vanschoren, J., García-Marín, A.P., 2022. AgroML: An Open-Source Repository to Forecast Reference Evapotranspiration in Different Geo-Climatic Conditions Using Machine Learning and Transformer-Based Models. Agronomy 12. https://doi.org/10.3390/agronomy12030656

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. Journal of Irrigation and Drainage Engineering 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. 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. https://doi.org/10.1002/eco.70157

Gao, R., 2025. met_gfit: A Python toolkit for comprehensive model evaluation and diagnostic plotting. Zenodo. https://doi.org/https://doi.org/10.5281/zenodo.18807443

Gao, R., Alsina, M.M., Torres-Rua, A., Hipps, L., Kustas, W.P., Anderson, M., Nieto, H., McElrone, A.J., Knipper, K., Bambach, 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., 2025a. Integrating Time-Series Meteorological Data and sUAS Information into a Machine Learning Framework for California Vineyard Water Stress Monitoring. https://doi.org/10.21203/rs.3.rs-7952103/v1

Gao, R., Alsina, M.M., Torres-Rua, A.F., Hipps, L., Kustas, W.P., White, W.A., Anderson, M., Alfieri, J., Dokoozlian, N., Nieto, H., Gao, F., McKee, L.G., Prueger, J.H., Sanchez, L., Mcelrone, A.J., Bambach Ortiz, N., Coopmans, C., Gowing, I., 2022a. Exploratory analysis of vineyard leaf water potential against UAS multispectral and temperature information, in: Thomasson, J.A., Torres-Rua, A.F. (Eds.), . SPIE, pp. 160–185. https://doi.org/10.1117/12.2622995

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

Gao, R., Torres-Rua, A.F., Aboutalebi, M., White, W.A., Anderson, M., Kustas, W.P., Agam, N., Alsina, M.M., Alfieri, J., Hipps, L., Dokoozlian, N., Nieto, H., Gao, F., McKee, L.G., Prueger, J.H., Sanchez, L., Mcelrone, A.J., Bambach-Ortiz, N., Coopmans, C., Gowing, I., 2022b. LAI estimation across California vineyards using sUAS multi-seasonal

multi-spectral, thermal, and elevation information and machine learning. Irrig. Sci. 40, 731–759. https://doi.org/10.1007/s00271-022-00776-0

Gao, R., Torres-Rua, A.F., Hipps, L., Kustas, W.P., Anderson, M., White, W.A., Alfieri, J.G., Alsina, M.M., Dokoozlian, N., Nieto, H., Gao, F., McKee, L.G., Prueger, J.H., Sanchez, L., Mcelrone, A.J., Bambach-Ortiz, N., Coopmans, C., Gowing, I., 2022c. Assessment of TSEB-PT and -2T in ET partitioning estimation over California commercial vineyards based on sUAS information. SPIE, p. 121140I. https://doi.org/10.1117/12.2622993

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. (Basel). 15, 756. https://doi.org/10.3390/rs15030756

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. Science of the Total Environment 957. https://doi.org/10.1016/j.scitotenv.2024.177606

Gyaneshwar, A., Mishra, A., Chadha, U., Raj Vincent, P.M.D., Rajinikanth, V., Pattukandan Ganapathy, G., Srinivasan, K., 2023. A Contemporary Review on Deep Learning Models for Drought Prediction. Sustainability (Switzerland). https://doi.org/10.3390/su15076160

Jha, P.K., Pathak, T.B., 2024. Seasonal climate forecasts show skill in predicting winter chill for specialty crops in California. Communications Earth & Environment 2024 5:1 5, 1–13. 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. https://doi.org/10.1016/j.agwat.2024.109219

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. 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. https://doi.org/10.1061/jhyeff.heeng-6161

Melton, F.S., Huntington, J., Grimm, R., Herring, J., Hall, M., Rollison, D., Erickson, T., Allen, R., Anderson, M., Fisher, J.B., Kilic, A., Senay, G.B., Volk, J., Hain, C., Johnson, L., Ruhoff, A., Blankenau, P., Bromley, M., Carrara, W., Daudert, B., Doherty, C., Dunkerly, C., Friedrichs, M., Guzman, A., Halverson, G., Hansen, J., Harding, J., Kang, Y., Ketchum, D., Minor, B., Morton, C., Ortega-Salazar, S., Ott, T., Ozdogan, M., ReVelle, P.M., Schull, M., Wang, C., Yang, Y., Anderson, R.G., 2022. OpenET: Filling a Critical Data Gap in Water Management for the Western United States. JAWRA Journal of the American Water Resources Association 58, 971–994. https://doi.org/10.1111/1752-1688.12956

Peddinti, S.R., Nicolas, F., Raij-Hoffman, I., Kisekka, I., 2025. Measurement and Modeling of Evapotranspiration for Processing Tomatoes Under Subsurface Drip Irrigation Using Eddy Covariance, pyfao56, and Particle Swarm Optimization. Journal of Natural Resources and Agricultural Ecosystems 3, 149–163. https://doi.org/10.13031/jnrae.16377

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 (Switzerland) 12. https://doi.org/10.3390/AGRICULTURE12121971

Safre, A.L.S., Nassar, A., Torres-Rua, A., Aboutalebi, · Mayhar, João, ·, Saad, C.C., Manzione, R.L., Antonio, ·, De, H., Teixeira, C., Prueger, J.H., Lynn, ·, Mckee, G., Alfieri, J.G., Lawrence, ·, Hipps, E., Nieto, H., White, W.A., Del Mar Alsina, M., Luis Sanchez, ·, Kustas, W.P., Dokoozlian, · Nick, Feng Gao, ·, Anderson, M.C., 2022. Performance of Sentinel-2 SAFER ET model for daily and seasonal estimation of grapevine water consumption. Irrigation Science 2022 1, 1–20. https://doi.org/10.1007/S00271-022-00810-1

Sarwar, A., Gao, R., Safeeq, M., Abatzoglou, J.T., Medellín-Azuara, J., Viers, J.H., 2025. Artificial Intelligence for Multi-Scale Drought Modeling and Decision Making, in: Fares, A. (Ed.), Thirsting for Change: Confronting the Global Drought Challenge with Innovative Solutions and Resilience Strategies. Elsevier.

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 (Weinh). 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

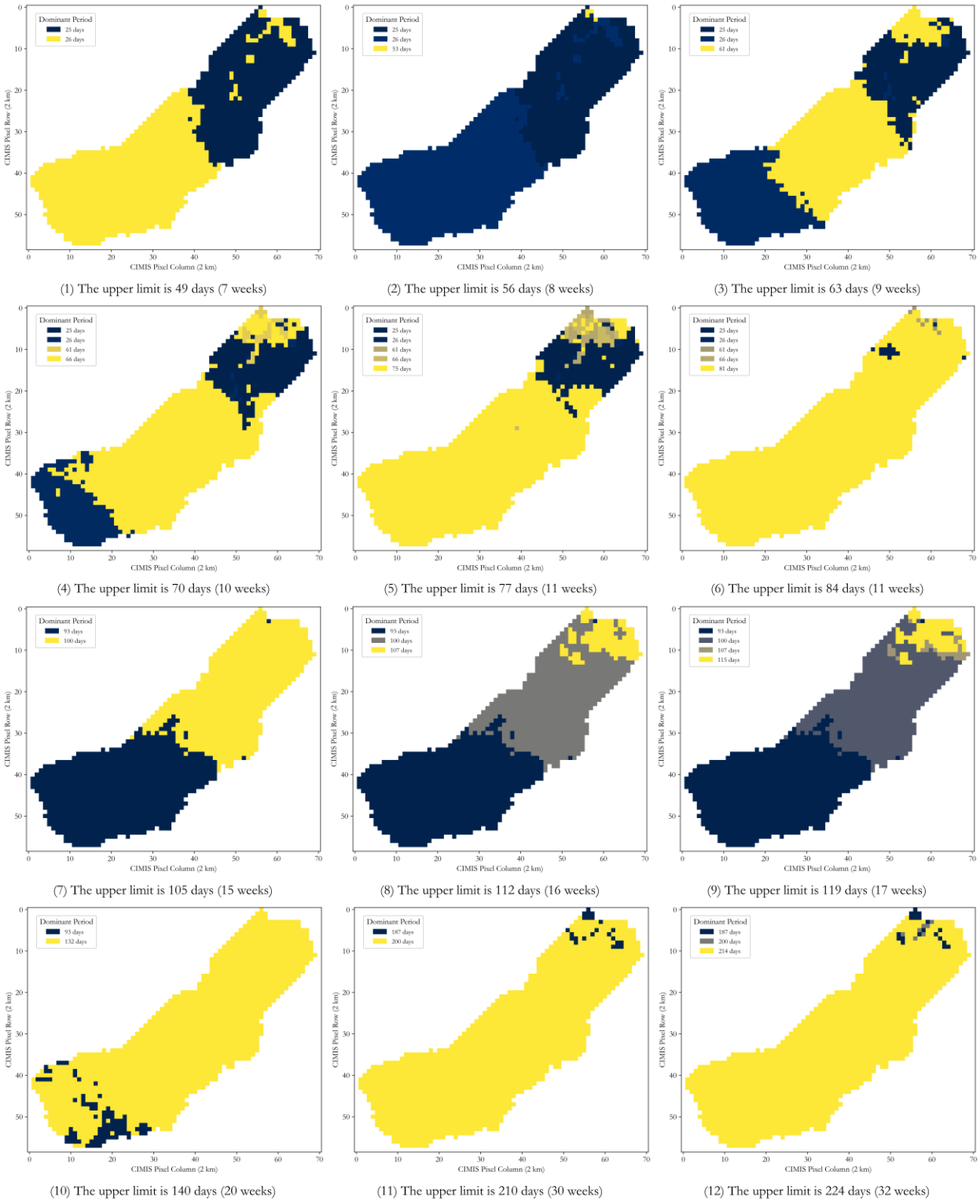
Sorooshian, S., Li, J., Hsu, K.L., Gao, X., 2011. How significant is the impact of irrigation on the local hydroclimate in Californias Central Valley? Comparison of model results with ground and remote-sensing data. Journal of Geophysical Research Atmospheres 116. https://doi.org/10.1029/2010JD014775

Srinivasan, J., Bernacchi, L.A., Eriksson, M., Safeeq, M., Waring, E., Viers, J.H., 2025. Assessing climate and water information systems affordances for the agricultural sector in western United States. PLOS Water 4. https://doi.org/10.1371/journal.pwat.0000458

Volk, J.M., Huntington, J.L., Melton, F.S., Allen, R., Anderson, M., Fisher, J.B., Kilic, A., Ruhoff, A., Senay, G.B., Minor, B., Morton, C., Ott, T., Johnson, L., Comini de Andrade, B., Carrara, W., Doherty, C.T., Dunkerly, C., Friedrichs, M., Guzman, A., Hain, C., Halverson, G., Kang, Y., Knipper, K., Laipelt, L., Ortega-Salazar, S., Pearson, C., Parrish, G.E.L., Purdy, A., ReVelle, P., Wang, T., Yang, Y., 2024. Assessing the accuracy of OpenET satellite-based evapotranspiration data to support water resource and land management applications. Nature Water 2, 193–205. https://doi.org/10.1038/s44221-023-00181-7

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. https://doi.org/10.1016/j.ecolind.2024.112810

Appendix



(1) The upper limit is 49 days (7 weeks)

(2) The upper limit is 56 days (8 weeks)

(3) The upper limit is 63 days (9 weeks)

(4) The upper limit is 70 days (10 weeks)

(5) The upper limit is 77 days (11 weeks)

(6) The upper limit is 84 days (11 weeks)

(7) The upper limit is 105 days (15 weeks)

(8) The upper limit is 112 days (16 weeks)

(9) The upper limit is 119 days (17 weeks)

(10) The upper limit is 140 days (20 weeks)

(11) The upper limit is 210 days (30 weeks)

(12) The upper limit is 224 days (32 weeks)

Figure A 1. CWT results with different upper limits.