Journal of Hydrology: Regional Studies 65 (2026) 103324

Contents lists available at ScienceDirect

Journal of Hydrology: Regional Studies

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

Beyond VPD: Multi-sensor, sub-hourly microclimate improves machine-learning prediction of greenhouse evaporation in a Mediterranean semi-arid greenhouse

Nir Averbuch, Menachem Moshelion *

Institute of Plant Sciences and Genetics in Agriculture, The Robert H. Smith Faculty of Agriculture, Food and Environment, The Hebrew University of Jerusalem, Rehovot 7610001, Israel

ARTICLE INFO

Keywords:
Pan evaporation
Microclimate heterogeneity
Hysteresis
Machine learning
Evapotranspiration modeling

ABSTRACT

Study region: This study was conducted in a semi-arid Mediterranean region (Rehovot, Israel) within a semi-controlled greenhouse environment characterized by strong spatial microclimatic heterogeneity.

Study focus: Accurate estimation of evaporation under heterogeneous microclimates remains challenging, particularly at sub-daily timescales where steady-state formulation often fails. This study aimed to isolate purely physical evaporation dynamics and evaluate whether high-frequency, multi-sensor data combined with machine-learning approaches improve sub-daily evaporation prediction compared to conventional FAO-56-Penman–Monteith. Open-water evaporation was continuously measured using 62 high-throughput gravimetric pans, paired with 12 spatially distributed meteorological stations sampled every 3 min over one year (~2.4 million records).

New hydrological insights for the region: Spatial heterogeneity reached a temperature gradient of 8 °C and relative humidity differences of 16%. Strong hysteresis was observed, whereby identical atmospheric drivers produced different evaporation rates depending on the time of day. Temporal encoding (hour of day and day of year) and light intensity dominated predictive importance, while vapor pressure deficit contributed minimally. XGBoost achieved the highest accuracy (R^2 = 0.935) and remained robust with 10% of the dataset (R^2 = 0.935). Tree-based models consistently outperformed Neural-Network, which showed higher sensitivity to data reduction. FAO-56 Penman–Monteith showed substantially lower performance (R^2 = 0.637).Together, these results show that hysteresis and temporal structure are critical for accurate high-frequency evaporation prediction under complex, non-steady microclimates, motivating machine-learning frameworks that can learn lagged, non-linear responses.

1. Introduction

Optimal irrigation is crucial for maintaining high water balance and thus supporting crop yield maximization, as demonstrated by the direct relationship between irrigation quantity and crop yield (Liao et al., 2008). Manual physiological measurements, such as leaf gas exchange and water potential, provide accurate momentary data on leaf water status. Nevertheless, these methods are limited by

* Corresponding author.
E-mail address: menachem.moshelion@mail.huji.ac.il (M. Moshelion).

https://doi.org/10.1016/j.ejrh.2026.103324
Received 29 September 2025; Received in revised form 2 March 2026; Accepted 4 March 2026
Available online 17 March 2026

low throughput and their reliance on steady-state assumptions for each measurement and by limited spatial–temporal resolution. Moreover, they typically assess only parts of the plant, offering relative indicators of water status (e.g., leaf stomatal conductance or water potential) without capturing whole-plant water consumption. These limitations restrict our ability to monitor whole-plant dynamic responses to environmental fluctuations and constrain the practical implementation of effective irrigation strategies. Hence, accurately quantifying the absolute amount of whole-plant transpiration and soil evaporation (together known as evapotranspiration, ET_o) is necessary to determine appropriate irrigation practices and thereby to sustain crop yields (Van Dijk et al., 2021).

Conceptually, ET_o integrates the combined influence of two distinct water fluxes: evaporation and transpiration. Evaporation is primarily a passive physical process controlled by the atmospheric demand, surface moisture availability, and the surface energy balance. In contrast, transpiration is a biologically mediated flux regulated by plant stomatal conductance, linking water loss to physiological responses. Although both processes contribute to ET0, they differ in their sensitivity to environmental drivers and often operate across different temporal scales, making precise measurement and prediction challenging (Markulj Kulundži et al., 2021). Methods for estimating ET_o include indirect models that use meteorological data and direct measurements of plant and/or soil water loss. However, each approach has limitations, particularly in capturing fine-scale spatial–temporal variability. A widely used indirect method for estimating ET_o is the FAO-56 Penman-Monteith equation (FPME), estimates ET_o based on meteorological parameters such as temperature, relative humidity (RH), wind speed, and solar radiation (Allen and Pereira, 1998; Ghiat et al., 2021). Although effective at large scales, FPME relies on sparse weather station measurements averaged over time, potentially overlooking microclimatic variations significantly impacting water loss (Cui et al., 2016). Moreover, the FPME requires empirical crop coefficients to estimate plant transpiration. However, while transpiration is biologically regulated via active stomatal responses to environmental signals, soil evaporation is a passive process. Consequently, identical environmental conditions can induce fundamentally different responses in soil evaporation and plant transpiration. For instance, a sharp increase in atmospheric vapor pressure deficit (VPD) proportionally increases soil evaporation, whereas plants actively respond by closing stomata (Yaaran et al., 2019), resulting in a different transpiration vs. evaporation response pattern.

Given that transpiration constitutes a substantially larger fraction of evapotranspiration (Chen et al., 2019; Schlesinger and Jasechko, 2014; Wei et al., 2017; Yao et al., 2023), separating and understanding these distinct processes is critical for accurate ET_o prediction models and practical irrigation regimes, particularly in complex environmental system that require multivariate analytical approaches(Das, 2022). Detecting their dynamic behavior requires continuous measurement of the soil–plant–atmosphere continuum to distinguish biologically regulated plant responses from passive environmental forcing and measurement noise arising from device limitations, errors, and hysteresis.

Our goal is to quantify plant transpiration, yet achieving this objective requires disentangling the complex, multifactorial influences of the environment on ET0 water loss. Due to the complexity of addressing these multifactorial challenges simultaneously, we chose to examine them in a stepwise manner. In this study, we focus on how environmental fluctuations affect evaporation, integrating high-resolution sensing with a rigorous data science workflow, including quality-controlled data acquisition, feature engineering, and statistical and machine-learning (ML) analyses, to characterize the microclimatic drivers of plant water use.

Seasonal variations, transient weather events, and the precise spatial positioning of evaporation pans within the greenhouse generate strong microclimatic heterogeneity due to differential exposure to radiation and airflow (Cabrera-Bosquet et al., 2016). These drivers often induce hysteretic evaporation responses, whereby evaporation does not follow environmental forcing instantaneously, complicating interpretation based on steady-state assumptions. Such dynamics impose fundamental limitations on conventional empirical and physically based models, which rely on predefined functional relationships and simplified temporal representations (Das, 2025; Shamshiri et al., 2018).

ML approaches offer a fundamentally different modeling paradigm by learning complex, nonlinear relationships directly from high-frequency, multivariate data without relying on explicit parametric formulations. This enables ML models to capture temporal dependencies, delayed system responses, and spatial heterogeneity that are difficult to represent within traditional frameworks. However, many existing ML-based ET_o studies rely on daily averaged or spatially aggregated inputs, which inherently limit their ability to resolve short-term microclimatic variability and hysteresis, particularly in heterogeneous environments (Balasubramanian and Thirugnanam, 2023; Başağaoğlu et al., 2021; Das, 2025, 2024; de Meneses et al., 2020; Ferreira et al., 2019; Kim et al., 2022). Additional methodological limitations include training-validation strategies that confound temporal generalization and the use of clustering approaches that may overfit narrow environmental conditions. Addressing these limitations requires high-resolution, spatially distributed environmental measurements acquired continuously over extended periods. In today's data-driven era, improved prediction hinges on the systematic collection of accurate, multivariate in situ measurements across heterogeneous sites at high temporal resolution. Our study directly addresses this need and aims to test whether fine-scale meteorological data (free from biological confounding) can better capture evaporation dynamics. To that end, we hypothesize that high-resolution meteorological measurements, recorded simultaneously with pan water loss rates, will provide a more accurate representation of evaporation dynamics than conventional coarse-resolution data. Furthermore, we hypothesize that machine-learning models trained on such high-resolution datasets will yield improved evaporation estimates relative to traditional empirical methods. To test these hypotheses, we utilize detailed environmental and pan evaporation measurements collected continuously, at three-minute intervals from 72 spatially distributed locations over an entire year. These data encompass continuous fluctuations in temperature, relative humidity, VPD, wind speed, and solar radiation. By explicitly excluding plant-driven influences, our study isolates environmental factors governing evaporation, thereby ensuring that observed variations in evaporation rates reflect solely microclimatic conditions, free from biological noise associated with transpiration. By isolating environmental variability, this study aims to establish a baseline that may facilitate future integration of biologically regulated transpiration.

2. Materials and methods

2.1. Experimental design and study site

The experiment was conducted over a full year (December 2020–2021) in a semi-controlled greenhouse located at the I-CORE Center for Functional Phenotyping (31.904167° N, 34.800654° E). The structure measured 18 × 16 m, with a gutter height of 4.5 m, a ridge height of 6 m, and estimated internal volume of 1417.5 m^3. The greenhouse was exposed to natural environmental variability, with the climate control system operating under defined thresholds: evaporative cooling pads were activated when air temperature exceeded 31℃, while heaters were activated below 8℃, and four circulation fans (18,000 $m^3\,h^{-1}$; total 72,000 $m^3\,h^{-1}$). Fans operated automatically from 30 min before sunrise until 30 min after sunset to maintain a stable, low-velocity airflow and minimize variability arising from uncontrolled wind effects (see Suppl Fig. 3). Temperature thresholds for heating and cooling activation were controlled by the greenhouse climate system with a sensor resolution of ±0.5 °C, and system states (heating, cooling, ventilation) were logged continuously at the same temporal resolution as the environmental measurements (3-min intervals). The experimental area was kept free of vegetation, ensuring that all recorded water loss reflected atmospheric-driven evaporation only, without biological contributions from plant transpiration and shading.

2.2. Gravimetric monitoring platform

Evaporation was measured using a telemetric gravimetric system (Plantarray; Plant-DiTech, Israel), which recorded pan weight at 3-minute intervals. The system included 72 load-cell lysimeters, 62 of which were mounted with evaporative pans (22 cm diameter, 5 cm depth) and automatically refilled daily at 04:00 to maintain a consistent initial water level. The system was mounted on an experimental bench at the greenhouse (7 m × 1.2 m) arranged in four rows (a–d) and 18 columns (1–18; Fig. 1). The other 10 lysimeters carried constant weights and served as reference controls for long-term drift correction and sensor validation. Observed load-cell drift remained below 15 g (range 2.7–5.3 g; min-max), corresponding to < 0.6% (range: 0.1–0.57%) of their initial weight (see Suppl Table 1 for full summary statistics).

2.3. Environmental Sensing and Spatial Interpolation

Environmental parameters were monitored using a distributed network of sensors integrated with the PlantArray system. Twelve VP-4 (METER Group, USA) meteorological stations (MS) measured air temperature, relative humidity (RH), and barometric pressure from which vapor pressure deficit (VPD) was calculated. To capture microclimatic variability, 12 MS units strategically placed at two vertical positions representing distinct aerodynamic layers above the bench (Z0 = 0.3 m, Z1 = 1.5 m corresponding to approximately 0.9 m and 2.1 m above the greenhouse floor, respectively) across six evenly spaced horizontal locations along the bench. The upper

Fig. 1. Experimental greenhouse bench equipped with a gravimetric evaporation platform and spatially distributed meteorological sensors. The experimental bench (7 m × 1.2 m × 0.9 m) was arranged in 18 columns (1–18) and 4 rows (a–d) and included 72 load-cell lysimeters (Plant-DiTech, Israel). Sixty-two lysimeters were mounted with evaporation pans (22 cm diameter, 5 cm depth), automatically refilled daily at 04:00. Ten additional lysimeters held constant weights for drift correction. Twelve meteorological stations (MS), measuring temperature, relative humidity, and barometric pressure, were installed at six positions: (1,a), (1,d), (9,a), (9,d), (18,a), and (18,d), each with two vertical heights: Z_0 (30 cm above pan level, marked with green arrow) and Z_1 (1.5 m, marked with orange arrow). The main meteorological mast (MMS), located near (18, d), recorded additional environmental parameters.

level (approximately 2.1 m above the floor) approximates the reference meteorological height used for FAO-56 Penman–Monteith ET_o estimation, whereas the lower level (approximately 0.9 m above the floor) captures the near-surface layer directly above the evaporative pans (see Fig. 1). The water temperature of 32 evaporative pans was measured using 5TE sensors. In addition, a WatchDog 2000 Series weather station (Spectrum Technologies, USA) recorded solar radiation, and an ATMOS 22 ultrasonic anemometer recorded wind speed. To estimate meteorological conditions above each lysimeter, we applied inverse distance weighting (IDW) interpolation (Eq. (1)) using data from the three nearest VP-4 meteorological stations. Interpolation accuracy analysis demonstrated that using three nearby stations was sufficient to reconstruct meteorological parameters with high accuracy ($p < 0.001$; Suppl. Fig. 1). The interpolation is defined as:

Eq. (1). Inverse-distance weighting (IDW) interpolation

$$Z_p = \frac{\sum_{i=1}^{n}\left(\frac{Z_i}{d_i^p}\right)}{\sum_{i=1}^{n}\left(\frac{1}{d_i^p}\right)}$$

D_i = distance between location i and the prediction points p, Z_i= observed values at location, Z_p = interpolated values at a location.

2.4. Penman-Monteith reference evaporation estimation

Reference evaporation (ET_o) was calculated using the FAO-56 Penman–Monteith equation, which integrates multiple meteorological parameters to estimate the potential evaporation from a hypothetical reference surface. The equation was applied at both 3-min and daily time scales using the interpolated environmental data described above. No crop coefficient or plant-related variables were included as the experiment was designed to capture only the physical component of evaporation. The formula used was:

Eq. (2). FAO-56 Penman-Monteith equation.

$$ET_o = \frac{0.408 * \Delta(R_N - G) + \gamma * \frac{900}{T+273} * U_2(e_s - e_a)}{\Delta + \gamma * (1 + 0.34 * u_2)}$$

ETo is the reference evaporation ($\frac{mm}{time}$), Rn represents the net solar radiation ($\frac{MJ}{m^2 * time}$), G represent soil heat flux ($\frac{MJ}{m^2 * time}$), T represents temperature (℃), U2 represents wind speed ($\frac{m}{time}$), (es − ea) represents the VPD (kPa), Δ represents the slope of the vapor pressure curve ($\frac{kPa}{℃}$), and γ represents the psychometric constant (kPa ℃$^{-1}$).

2.5. Data processing, quality control, and filtering

Prior to analysis, all gravimetric and meteorological time series data underwent a systematic data processing and quality control workflow to ensure internal consistency across sensors and time. Raw measurements were first resampled to uniform 3-minute temporal resolution, matching the gravimetric acquisition frequency. Sensor drift in the gravimetric measurements was assessed using reference lysimeters carrying constant weights, confirming minimal long-term drift (<0.6% of initial weight; see Suppl. Table 1). Missing or invalid data records, primarily resulting from short communication or logging interruptions, were identified using continuity and plausibility checks. Short gaps were linearly interpolated only when limited to brief durations (up to 9 min) and when required to maintain temporal continuity, whereas longer gaps were excluded from subsequent analyses. Physically implausible values and transient spikes were removed using sensor-specific interquartile range (IQR) filtering, applied independently to each time series. Any remaining missing values after filtering were excluded prior to statistical analysis and model training. All subsequent analyses were conducted on the quality-controlled dataset.

2.6. Machine learning models and training protocol

To model open evaporative pans using environmental variables, four supervised machine-learning algorithms were implemented: Decision Tree, Random Forest, Extreme Gradient Boosting (XGBoost), and Artificial Neural Network (ANN). Models were deployed in Python using the following libraries: **scikit-learn** for Decision Tree and Random Forest (Pedregosa et al., 2012), **XGBoost** for gradient-boosted trees (Chen and Guestrin, 2016), and **TensorFlow** for the ANN (Abadi et al., 2016).

Table 1
Hyperparameter configuration of the machine-learning models used for evaporation prediction.

Model	Key Hyperparameters
Decision Tree	min_samples_split = 2; no maximum depth
Random Forest	n_estimators = 100; min_samples_split = 2;no maximum depth
XGBoost	max_depth = 100; n_estimators = 55;learning_rate = 0.3; reg_lambda = 0.3; gamma = 1e-5
ANN	5 hidden layers; epochs = 100; batch size = 1024; optimizer = Adam; learning_rate = 1e-3

Summary of key hyperparameters for the Decision Tree, Random Forest, XGBoost, and Artificial Neural Network (ANN) models used in this study. Parameters were selected to allow flexible model structure and stable training across the high-resolution evaporation dataset.

Each model was trained to predict the evaporation rate at 3-minute resolution, using air temperature, RH, VPD, solar radiation, wind speed, pan water temperature, and temporal encodings (minute, hour of day, day of year). Hyperparameters were selected based on preliminary sensitivity analyses to balance model flexibility with stable convergence and reproducible performance across different datasets formats. Model-specific hyperparameters are summarized in Table 1. For the tree-based models (Decision Tree, Random Forest, and XGBoost), parameters were tuned to allow deep, flexible partitioning of the feature space without imposing restrictive depth or pruning constraints. The ANN was configured with five hidden layers and optimized using the Adam algorithm, with the number of training epochs (100) and batch size (1024) selected to ensure stable training behavior while avoiding overfitting.

The final data set (all dataset variations are listed in Suppl. Table 3) consisted of approximately 2.4 million records, recorded at 3-minute intervals from 62 lysimeters, excluding the 10 reference lysimeters used to monitor sensor drift. The dataset was randomly split into 70% for training and 30% for testing. To ensure consistent model exposure to identical train/test splits, a fixed random seed (42) was used for all models to ensure reproducibility, with additional seeds tested during selectivity analysis. Model performance was assessed using the coefficient of determination (R^2), mean absolute error (MAE) and root mean square error (RMSE) and confidence intervals. Statistical comparisons among models were conducted using one-way analysis of variance (ANOVA) in JMP Pro 16. Model performance was evaluated using complementary metrics: R^2 to quantify explained variance, MAE and RMSE to capture absolute and squared error magnitudes, and bootstrap-derived confidence intervals to assess robustness and uncertainty.

3. Results

3.1. Spatial–temporal microclimatic variation affects evaporation homogeneity

The first part of this study focused on quantifying the spatial and temporal variability of environmental parameters within the greenhouse. Specifically, we examined daily vapor pressure deficit (VPD) patterns along with their two primary components: temperature and relative humidity (RH). As shown in Fig. 2a–d, the daily VPD maintained its typical bell-shaped pattern throughout the year yet showed increasing midday peak values and growing spatial differences across meteorological stations, particularly during the warmer seasons. To identify the sources of this heterogeneity, we analyzed temperature (Fig. 2e–h) and RH (Fig. 2i–l) gradients between the peripheral station (18, d–Z_0) and the opposite end of the bench (1, d–Z_0 and Z_1). Temperature differences of up to 8 °C and RH differences of up to 16% were observed between these locations. Temperature variability during the warmer months was especially pronounced at the elevated sensor height ($Z_1 = 1.5$ m) compared to the lower height ($Z_0 = 0.3$ m), though horizontal spatial differences

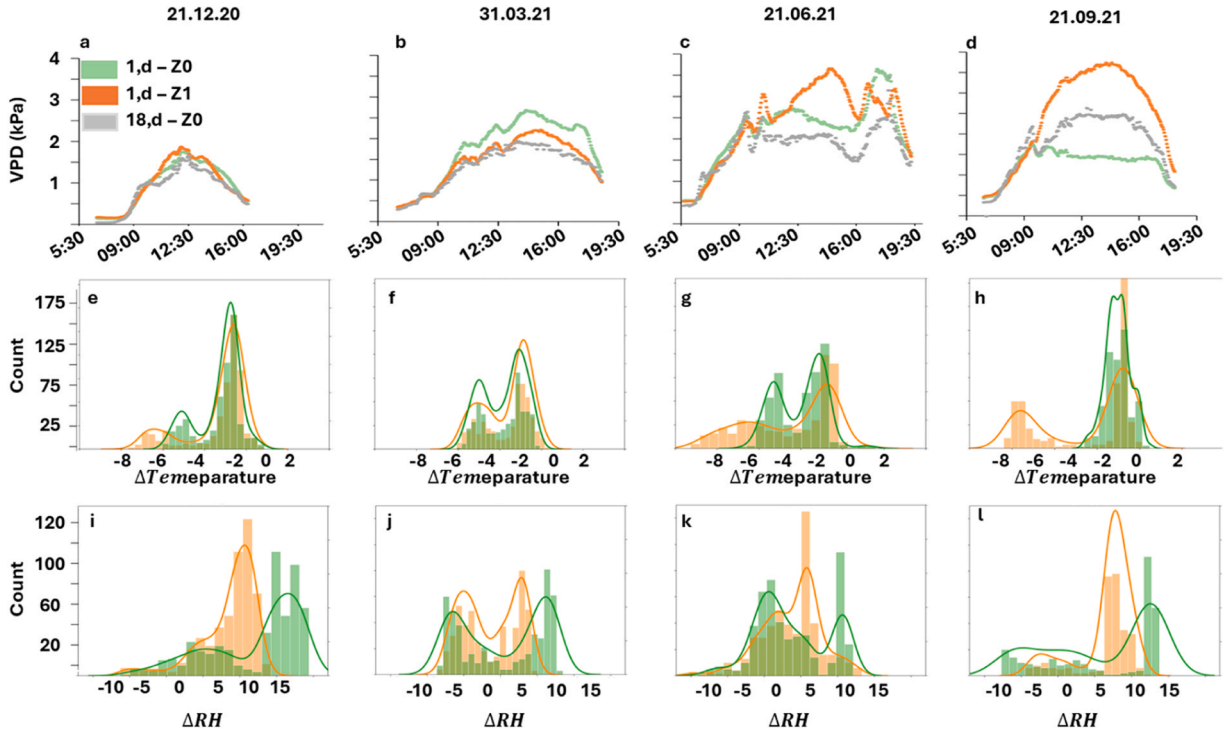


Fig. 2. Spatial-temporal variation in VPD and its components across the greenhouse bench. Continuous VPD measurements on four representative days (a–d), recorded at three meteorological stations: 18,d–Z_0 (gray line), located at one edge of the bench; 1,d–Z_0 (green); and 1,d–Z_1 (orange), located at the opposite edge at two different heights (see Fig. 1). Distribution of temperature (e–h) and distribution of relative humidity (i–l) for the same days, capturing spatial gradients across the bench. The differences between the stations reflect the microclimatic disparity between the two edges of the experimental platform, illustrating how temperature and humidity dynamics contribute to spatial heterogeneity in VPD.

between the bench edges remained dominant. These spatial gradients indicate substantial microclimatic heterogeneity within the greenhouse, suggesting that each evaporative pan was exposed to a distinct local environment. These findings motivated the interpolation of environmental data at each lysimeter each level, to better capture local conditions affecting evaporation dynamics. Thus, we interpolated environmental data using inverse distance weighing (IDW Eq. (1), Suppl. Fig. 1) to reconstruct high-resolution spatial patterns over time above each of the evaporative-pan lysimeters. The resulting projections illustrate the four-dimensional heterogeneity (space (x, y, z) X time) in temperature distribution across the greenhouse bench (Fig. 3).

Despite the pronounced differences in environmental conditions across the bench, no clear correlation was observed between midday VPD (measured at either Z_0 or Z_1; Fig. 4a–d) and the corresponding pan evaporation rates recorded by the lysimeters (Fig. 4e–h). Indeed, the relationship between evaporation and VPD consistently exhibited pronounced diurnal hysteresis patterns (Fig. 5), indicating a temporal decoupling between external atmospheric forcing and the evaporation response.

Evaporation rates and VPD both increase in magnitude across seasons, with lower values observed during winter (December) and progressively higher values during spring and summer, reaching maximum levels in late summer and early autumn (September) (Fig. 5a–b). The hysteresis loop between VPD–evaporation demonstrated that identical VPD values were associated with different evaporation rates in the morning versus the afternoon (black and gray respectively; Fig. 5c). indicating that evaporation does not follow instantaneous VPD alone, with predominantly clockwise but slightly different direction trajectories indicated by the arrows. Similar hysteretic relationships were observed between evaporation and additional environmental drivers, including air temperature, relative humidity (RH), and PAR, as shown in Suppl. Fig. 4.

3.2. *Microclimatic challenges for the FPME model*

Following the characterization of spatial and temporal heterogeneity across the greenhouse bench, we evaluated the accuracy of the FAO-56 Penman–Monteith equation (FPME) in estimating evaporation under these variable conditions. To enable spatial resolution, the model was applied using interpolated environmental inputs at each pan location and computed at both daily and 3-minute intervals. As shown in Fig. 6a, FPME estimates deviated from the lysimeter-measured daily evaporation by up to ±22%, depending on the season. In some cases, the model overestimated evaporation, while in others it underestimated it. Fig. 6b.1–b.4 illustrates that the calculated values exhibit a denser distribution, yet with greater deviation from the measured data's structure, indicating that FPME fails to fully capture localized microclimatic influences on evaporation. To test whether higher temporal resolution could improve FPME performance under spatially variable greenhouse conditions, we calculated FPME evaporation at 3-minute intervals. However, as shown in Fig. 7, the correlation between predicted and measured values remained moderate ($R^2 = 0.637$). These findings suggest

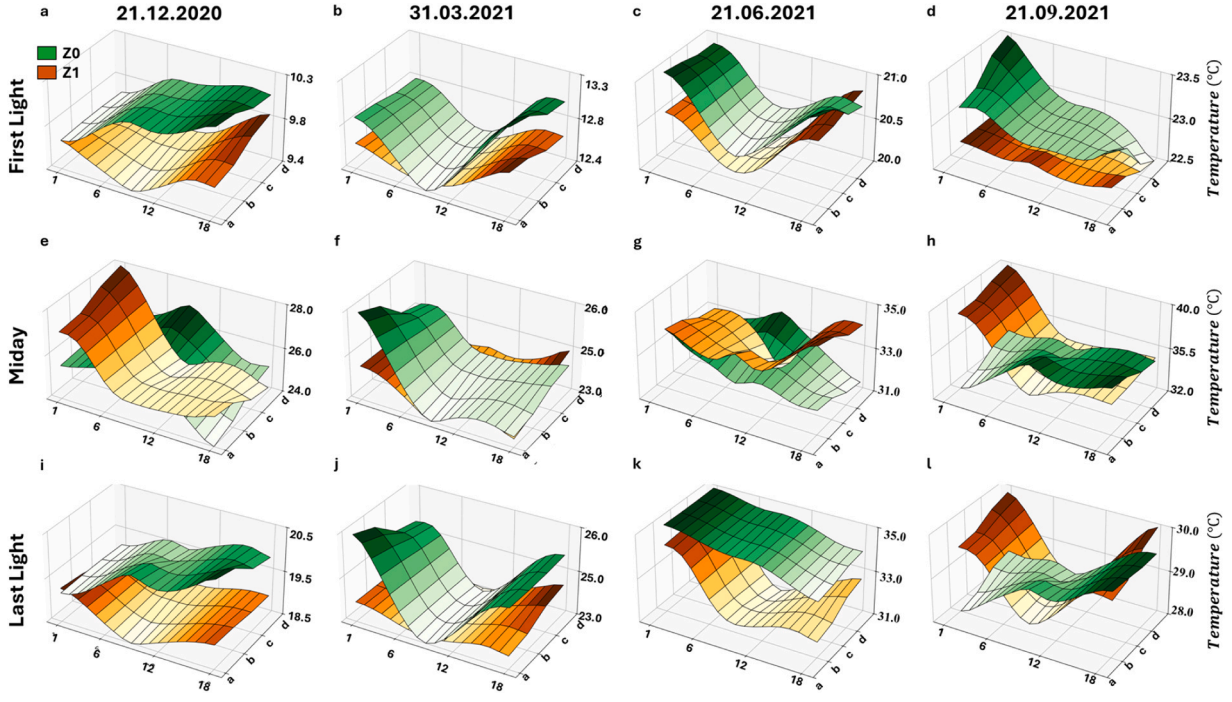


Fig. 3. Four-dimensional projection of spatial and temporal temperature variation across the greenhouse bench. Temperature maps above the 72 lysimeters at two heights, Z_0 (green) and Z_1 (orange), are shown for four representative dates and three time points: first solar radiation (a–d), midday (e–h; 12:00), and last solar radiation (i–l). Values at unmeasured locations were estimated using the inverse distance weighting (IDW) method based on the 12 meteorological stations (six at Z_0, six at Z_1; see Fig. 1 and Eq (1)). The temperature axis scale in each subplot is adjusted to its own minimum and maximum values. Distance between adjacent grid points is 0.3 m.

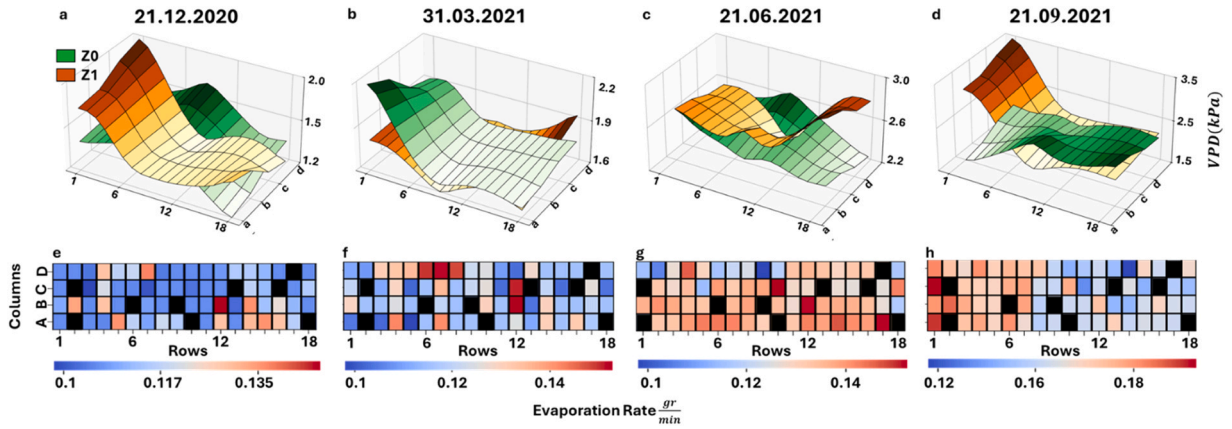


Fig. 4. Midday VPD and corresponding evaporation rates across the experimental bench. Spatial distribution of VPD measured at two heights (a–d), Z_0 (green) and Z_1 (orange), at four representative dates. Axis ranges are independently scaled according to the minimum and maximum VPD values in each subplot. Midday evaporation rates recorded by 62 lysimeter-mounted evaporation pans (e–h; see Materials and Methods). Black squares indicate positions of control lysimeters with fixed weights (non-evaporating).

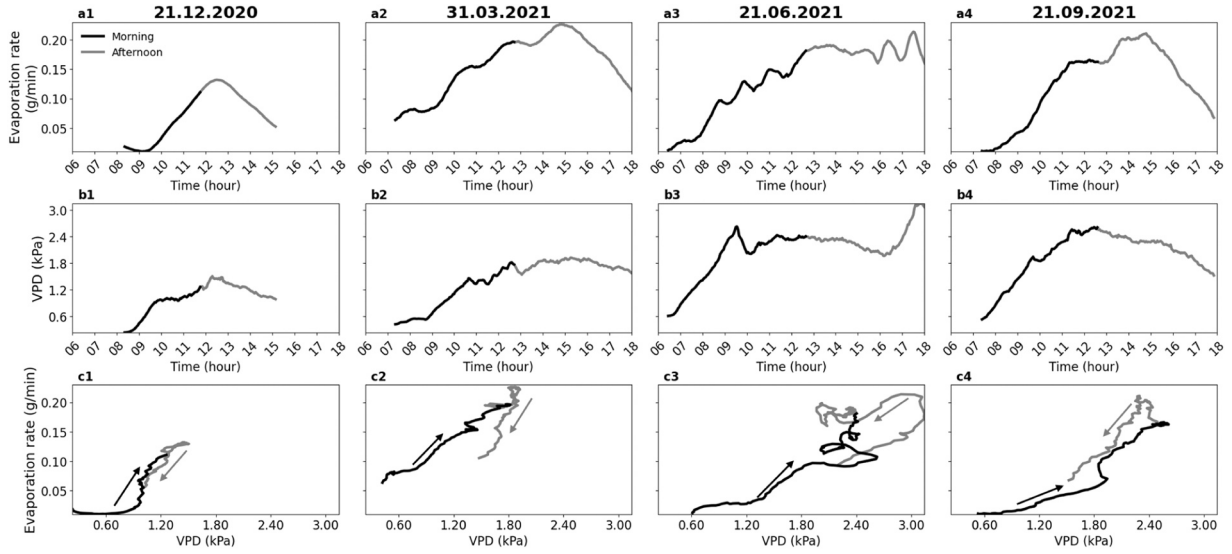


Fig. 5. Diurnal hystresis between vapor pressure deficit and evaporation rate across seasons. Diurnal evaporation rate measured at 3-minute interval resolution on four representative dates (a1–a4; 21.12.2020, 31.03.2021, 21.06.2021, 21.09.2021). Corresponsive diurnal pattern of vapor pressure deficit (VPD) for the same dates and time periods (b1–b4). Hystresis loops showing the relationship between evaporating and VPD over the course of the day (c1–c4). Black indicates morning hours (sunrise to solar noon), and gray indicates afternoon hours (solar noon to sunset). Arrow indicates the progression of time. Date shown for a representative evaporation pan, selected from the 62 lysimeters mounted pan.

that increasing temporal resolution alone is insufficient to account for the environmental heterogeneity captured by lysimeter measurements. Consequently, we next evaluated whether machine learning (ML) approaches could improve evaporation prediction under the same environmental input conditions.

3.3. *Predicting evaporation using machine-learning algorithms*

To evaluate whether higher spatial–temporal resolution improves predictive performance, we applied four machine learning (ML) algorithms: Decision Tree, Random Forest, XGBoost, and Artificial Neural Network, to estimate evaporation rates under heterogeneous greenhouse conditions (see Section 2, 2.4). Two input datasets were tested: (1) the indirect dataset, which included only ambient environmental variables (air temperature, RH, VPD, wind speed, and solar radiation), and (2) the indirect + direct dataset, which additionally included pan water temperature, assuming it integrates microclimatic effects and reflects the dominant physical driver of evaporation.

As shown in Fig. 8a–d, all models performed better with the inclusion of pan water temperature. Among them, XGBoost yielded the

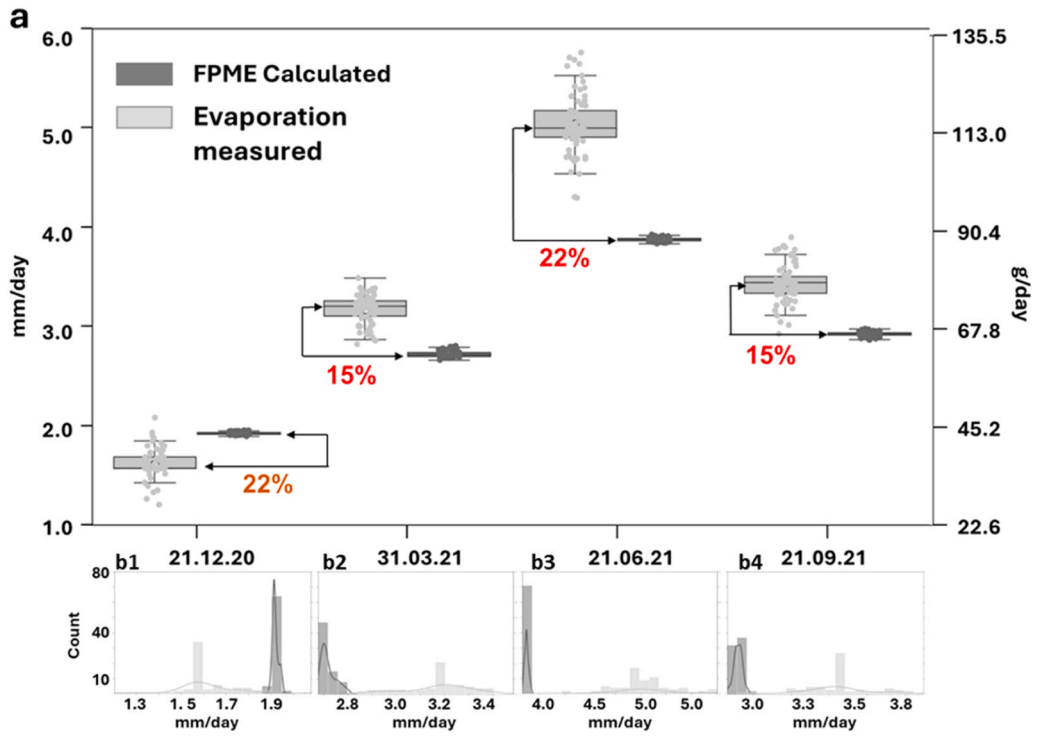


Fig. 6. Seasonal comparison between measured and FPME-calculated daily evaporation. Boxplots show daily evaporation rates across all lysimeters (n = 62) on four representative dates (a), comparing direct evaporation measurements (light gray) to FPME-calculated evaporation estimates (dark gray). Arrows indicate the percentage difference between medians: red for FPME overestimation, dark orange for underestimation. Distribution histograms of measured (light gray) and FPME-calculated (dark gray) evaporation values on the same dates, revealing seasonal variation in the accuracy and spread of the FPME estimates (b.1–b.4).

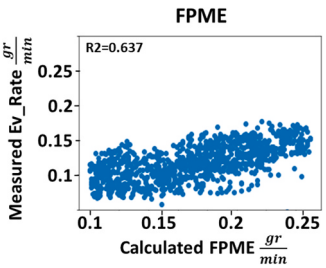
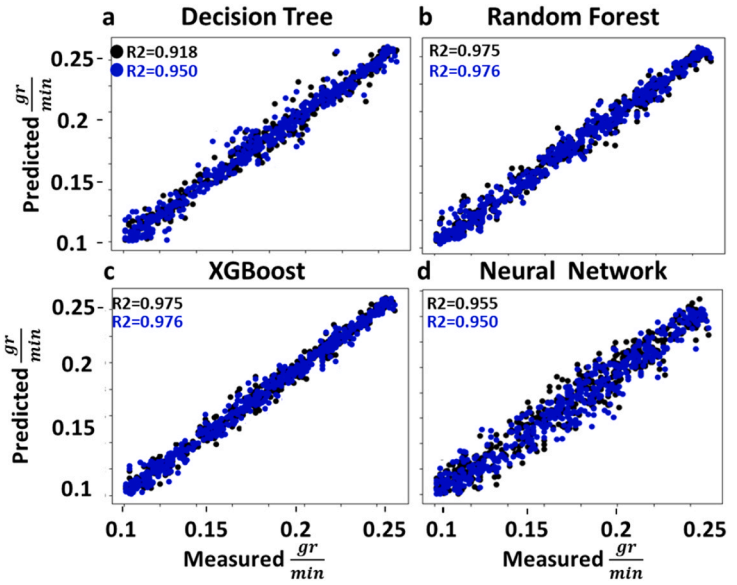


Fig. 7. Correlation between lysimeter-measured evaporation and FPME estimates at 3-min resolution. Scatterplot of 3-min evaporation rates comparing lysimeter-based measurements to values calculated using the FAO-56 Penman–Monteith equation (FPME) with spatially interpolated environmental inputs. The dataset includes over 3 million timepoints (n ≈ 3.01×10^6). Pearson correlation coefficient: $R^2 = 0.637$.

highest accuracy, achieving $R^2 = 0.976$ with the indirect + direct dataset (blue dots) bootstrap-based confidence interval for R^2 were (95% CI: 0.973–0.979 interval Suppl. Table 2) compared to $R^2 = 0.975$ and 0.973–0.977 confidence interval (Suppl. Table 2) using ambient variables alone (black dots), While the improvement was modest, it was consistent across models. An ANOVA-based comparison of absolute prediction error (Fig. 8e) confirmed that the inclusion of water temperature significantly improved model performance ($p < 0.001$). Across all models, confidence intervals (Suppl. Table 2) were narrow with relatively low MAE and RMSE, indicating that the observed ranking of the model accuracy was robust.

To quantify the relative influence of each input variable on evaporation prediction, we analyzed the importance hierarchies (Li et al., 2017) generated by all models under both input configurations (Fig. 9). Among the environmental variables, solar radiation consistently emerged as one of the most influential predictors, reflecting its central role in surface energy balance. In contrast, VPD received low importance across all models, despite its theoretical relevance to evaporation processes. Temporal features, particularly hour of day and day of year, were consistently ranked among the top three predictors in both input types, underscoring their strong predictive power. Notably in the XGBoost model (which achieved the highest accuracy), the hour of day was ranked as the most important feature. Although the inclusion of pan water temperature improved model performance, this variable often ranked below



e

	Indirect	Indirect + Direct
Decision Tree	A	A
Random Forest	B	A
XGBoost	B	B
Neural Network	C	C

Fig. 8. Comparison of machine learning model performance using indirect versus indirect + direct input datasets. A dataset comprising 1.2 million records of meteorological parameters and pan water temperature was used to predict evaporation rates. Each ML model was trained and tested separately using two input configurations: *Indirect* dataset (black dots): ambient environmental parameters only and *Indirect + Direct* dataset (blue dots): ambient parameters plus pan water temperature. Four models were evaluated on the same split dataset (70% training, 30% testing): Decision Tree (a), Random Forest (b), XGBoost (c), and Neural Network (d). Results of ANOVA comparing the absolute prediction error across models and input types (e). Bars labeled with different letters indicate statistically significant differences ($p < 0.001$) between models and between datasets (indirect vs. indirect + direct).

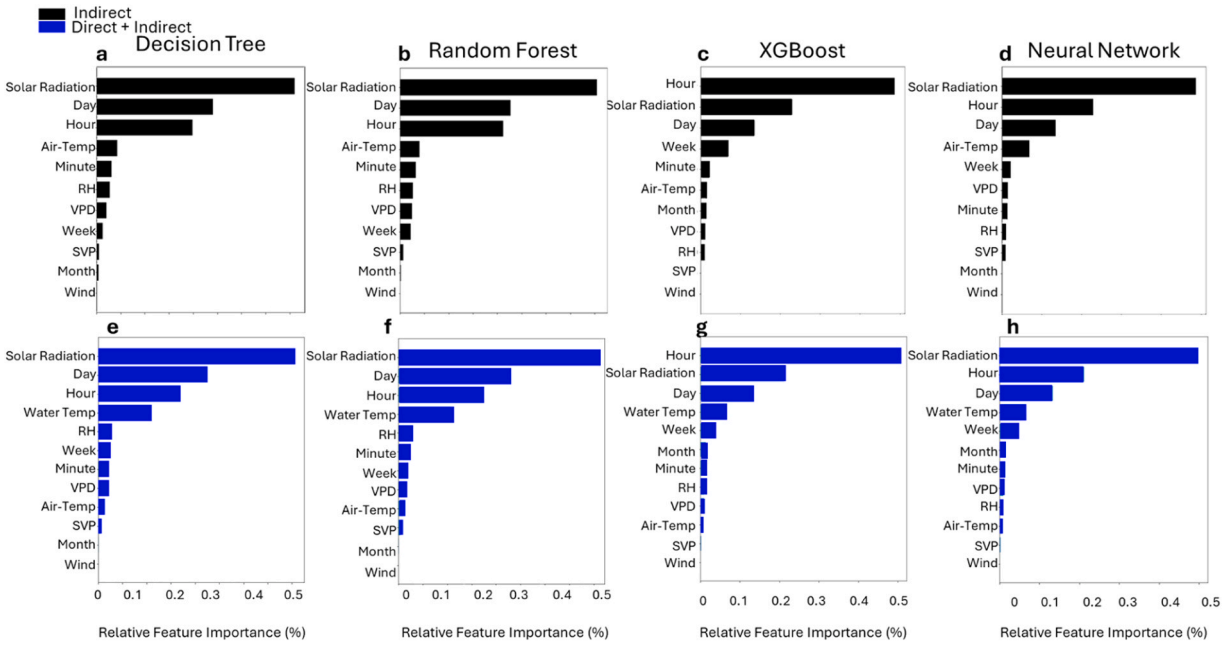
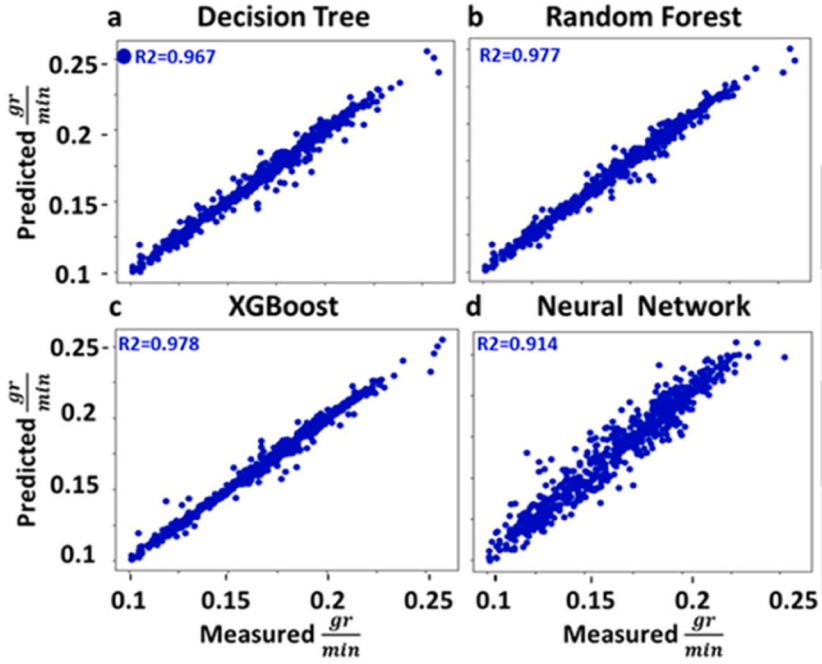


Fig. 9. Feature importance rankings across machine learning models trained on indirect and indirect + direct input datasets. Feature importance scores for four ML algorithms, were derived for two input types: *Indirect* (a–d, black bars) and *Indirect + Direct* (e–f, blue bars). Bar length indicates the relative contribution of each feature to model prediction. Temporal variables (e.g., hour of day, day of year) and solar radiation consistently ranked among the top predictors. The inclusion of pan water temperature altered feature rankings but did not override the dominance of key temporal features.

temporal features, even in the indirect + direct dataset, suggesting that temporal structure encapsulates key cumulative environmental effects.

To test whether accurate evaporation prediction could still be achieved with a minimal set of input features, we trained all models on a reduced dataset consisting only of pan water temperature as an independent physical variable, together with temporal encodings (hour of day and day of year). This analysis was designed to explicitly evaluate water as a standalone explanatory variable, rather than



e

	direct
Decision Tree	A
Random Forest	A
XGBoost	B
Neural Network	C

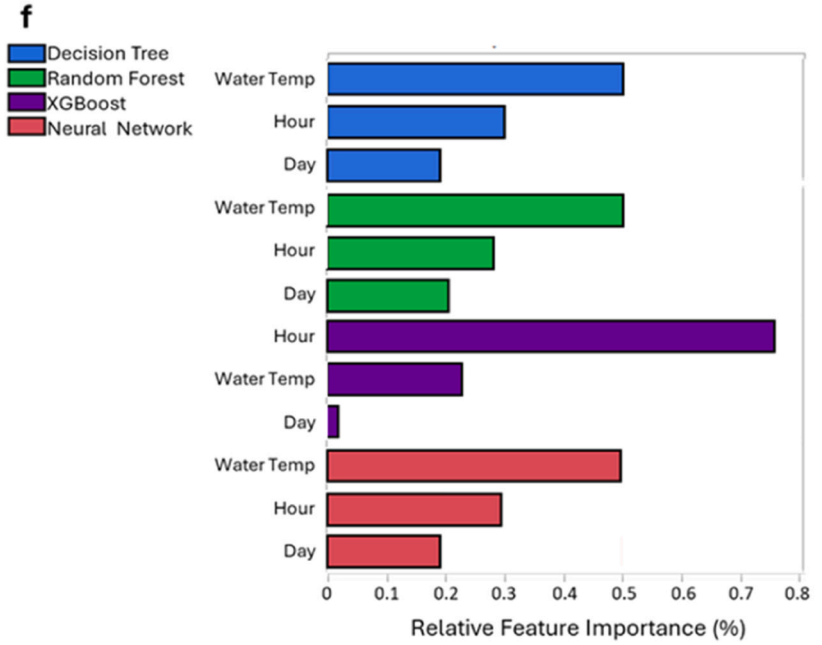


Fig. 10. Performance and feature importance of machine learning models trained on minimal input (water temperature and temporal data). A dataset of 1.2 million records containing pan water temperature and time-related variables was used to predict evaporation rates. All models were trained and tested using the same data split (70% training, 30% testing). Scatterplots comparing predicted versus measured evaporation for: Decision Tree(a), Random Forest (b), XGBoost (c), and Neural Network (d). ANOVA-based connecting letter report comparing the absolute prediction error across models (e). Different letters denote statistically significant differences ($p < 0.001$). Feature importance rankings within the reduced-input models (f). Colors represent model type: Decision Tree (blue), Random Forest (green), XGBoost (purple), and Neural Network (red).

as a proxy or alternative for ambient atmospheric parameters. As shown in Fig. 10a–d, the reduced-input models maintained high predictive performance, particularly for XGBoost and Random Forest. An ANOVA of absolute prediction error (Fig. 10e) confirmed a significant advantage for XGBoost over the other models ($p < 0.001$). The feature importance rankings of this model (Fig. 10f) showed that hour of day and day of year were consistently ranked above pan water temperature, even under this minimal configuration. To assess the practical feasibility of deploying these models under field constraints, we next evaluated their robustness across temporally and spatially limited scenarios.

3.4. Machine learning model robustness under temporal and spatial limitations

The accuracy of time-series predictions is influenced by both the temporal resolution of input data and the total volume of available observations. As data quantity increases, model performance is generally expected to improve (Es-Sabery et al., 2021). To evaluate the robustness of the tested models under practical constraints, we assessed their performance under varying degrees of temporal and spatial data reduction, including downsampled input, partial dataset coverage, and randomized observation subsets.

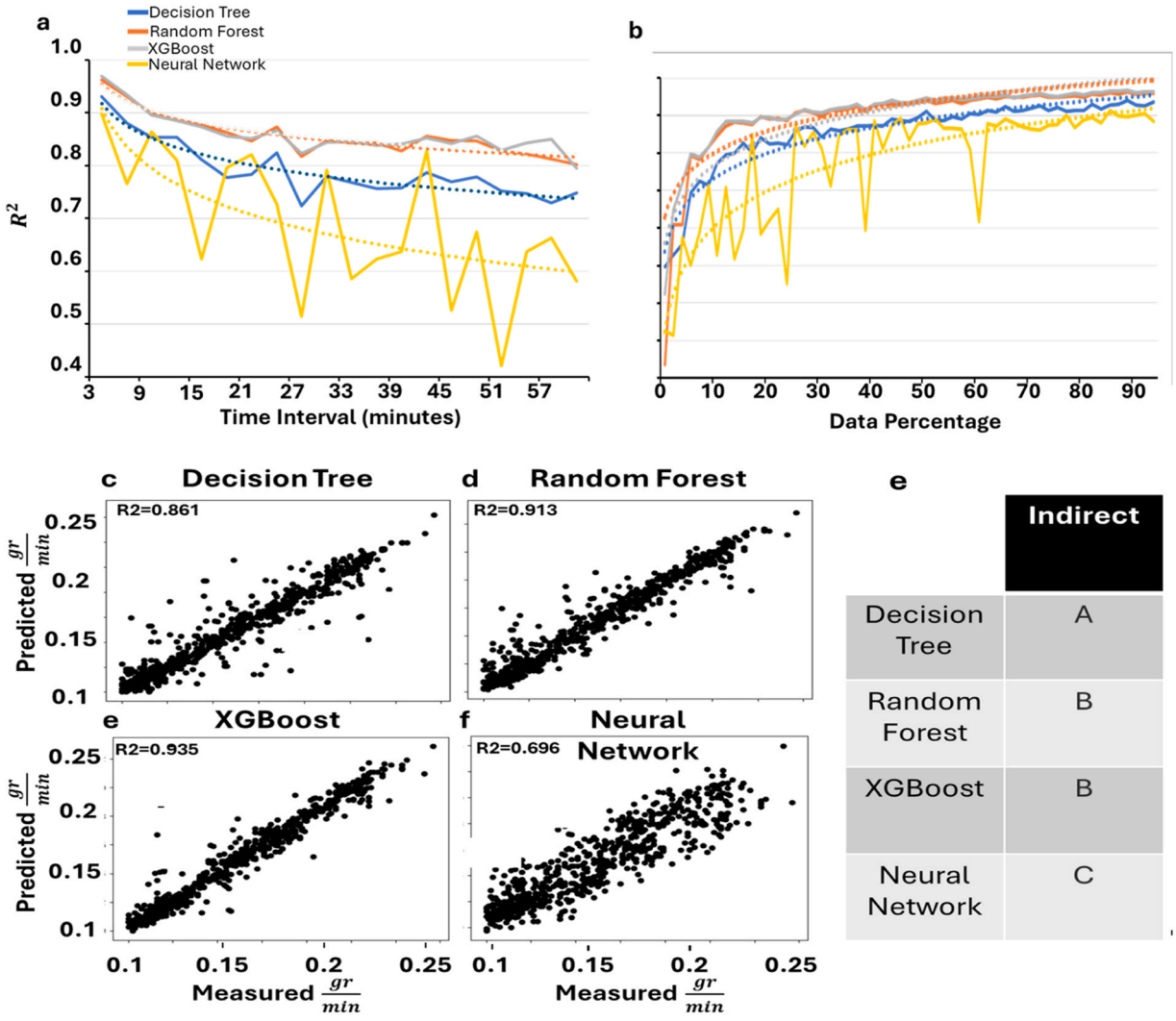
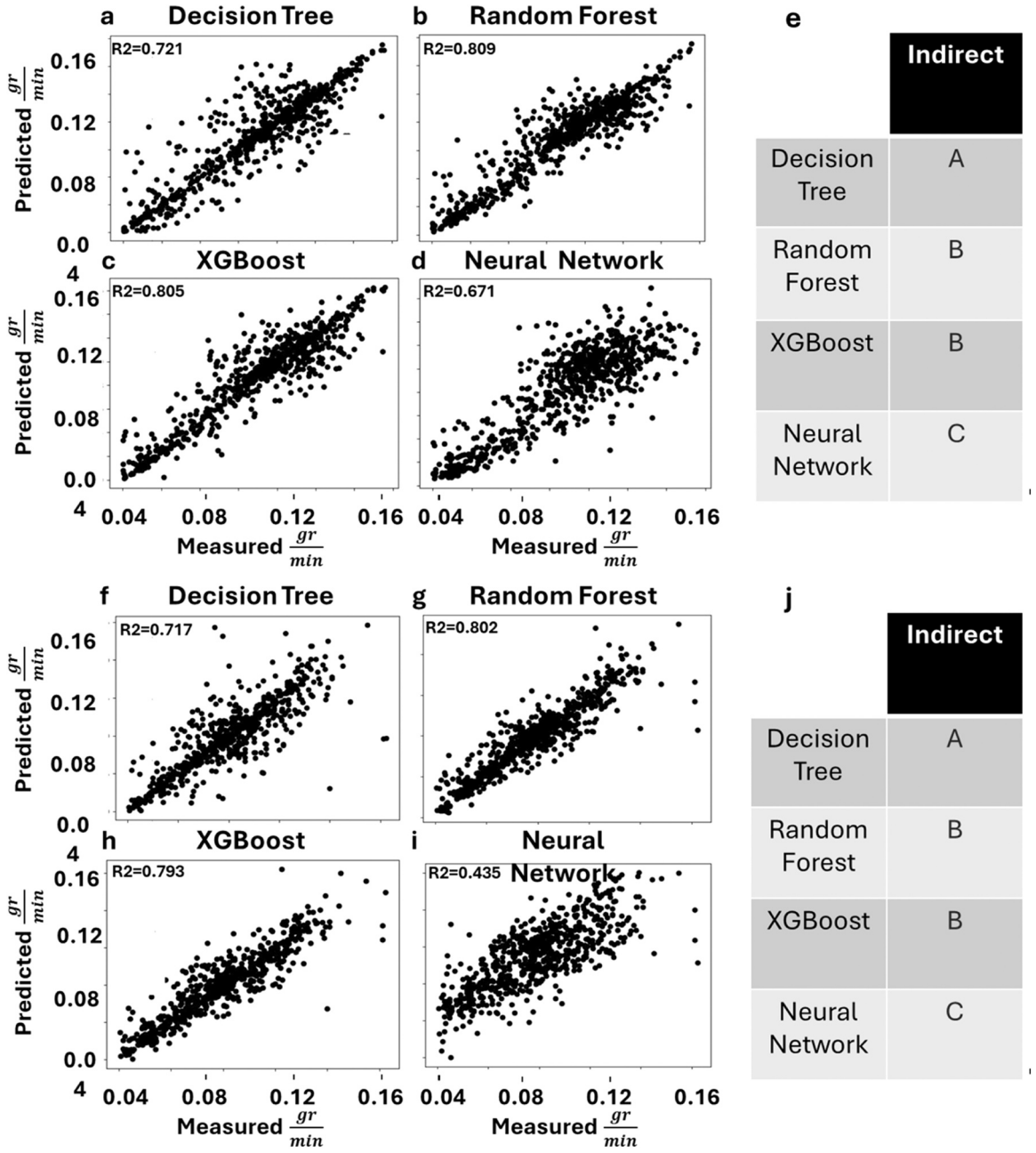


Fig. 11. Model robustness under temporal and spatial limitations. Prediction accuracy (R^2) of Decision Tree (blue), Random Forest (orange), XGBoost (gray), and Neural Network (yellow) models under varying data constraints: Temporal resolution (a): effect of increasing time intervals (3–60 min) on model performance logarithmic fit show R2 values of 0.835 to decision tree, 0.877 for random forest, 0.820 for XGBoost, and 0.418 for neural network. Training data volume (b): effect of training on increasing proportions of the dataset, logarithmic fit show R2 values of 0.925 to decision tree, 0.742 for random forest, 0.827 for XGBoost, and 0.714 for neural network. Scatterplots comparing predicted versus measured evaporation using only 10% of the temporal dataset (c–f; ambient parameters only) for: Decision Tree (c), Random Forest (d), XGBoost (e), Neural Network (f). ANOVA-based connecting letter report of absolute error for all models under this reduced-data condition (g). Different letters indicate statistically significant differences ($p < 0.001$).

Fig. 11a shows model performance when input data were aggregated into coarser time intervals. All models showed a performance decline with increasing sampling intervals. XGBoost, for example, dropped from $R^2 = 0.976$ at 3-min resolution to 0.928 at 30 min, and further to 0.857 at 60 min. In comparison, Decision Tree performance declined more steeply, reaching $R^2 = 0.705$ at 60-min intervals. Notably, the Neural Network model exhibited unstable performance, with fluctuating R^2 values across time intervals. Fig. 11b illustrates model performance as a function of the proportion of training data used. For all models, R^2 improved logarithmically with



e

	Indirect
Decision Tree	A
Random Forest	B
XGBoost	B
Neural Network	C

j

	Indirect
Decision Tree	A
Random Forest	B
XGBoost	B
Neural Network	C

Fig. 12. Spatial model transferability and minimal-data performance of machine learning algorithms. Machine learning models were evaluated in two experiments using meteorological data from an independent spatially distinct setup (three evaporation pans placed ~2.5 m apart): Models trained and tested solely on the small dataset (a–d; ~21,500 entries; ~0.715% of the original dataset maintained in the first experiment), using ambient (indirect) input parameters only. Decision Tree (a), Random Forest (b), XGBoost (c), Neural Network (d). ANOVA-based connecting letter report (e) for (a–d), comparing absolute prediction error. Different letters indicate statistically significant differences ($p < 0.001$). Generalization test (f–i): models trained on the full dataset (first experiment) plus the small-scale experiment and tested on a held-out portion of the small dataset without location labels: (f) Decision Tree, (g) Random Forest, (h) XGBoost, (i) Neural Network. (j) ANOVA comparison of absolute errors in (f–i); different letters denote significant differences ($p < 0.001$).

increasing data volume. Tree-based models (Random Forest, XGBoost, and Decision Tree) stabilized after 20–30% of the data, achieving R^2 values of 0.907, 0.908, and 0.867, respectively. In contrast, the Neural Network required over 60% of the data to stabilize, reaching $R^2 \approx 0.87$, and remained sensitive to subsampling variability.

This finding suggests that 20–30% of the original temporal dataset is sufficient to obtain $R^2 > 0.85$ prediction from all algorithms except the Neural Network. Testing only 10% of the full dataset ($n \approx 3.01 \times 10^5$) with ambient parameters only (excluding pan water temperature) accuracy declined across all models, yet XGBoost retained the highest performance ($R^2 = 0.935$), followed by Random Forest (Fig. 11c–f). ANOVA of the absolute prediction error (Fig. 11g) confirmed that XGBoost and Random Forest outperformed other models significantly ($p < 0.001$).

To evaluate model robustness under minimal spatial data conditions, we conducted an independent experiment on the opposite side of the greenhouse bench. This setup included only three evaporation pans spaced ~2.5 m apart, monitored over one month. The resulting dataset (~21,500 entries, ~0.715% of the main experiment) introduced significant spatial heterogeneity and limited data density. As shown in Fig. 12a–d, tree-based models maintained relatively high prediction accuracy under these constraints: Random Forest and XGBoost achieved R^2 values of 0.809 (95 CI: 0.795–0.818) and 0.805 (95 CI: 0.793–0.816), respectively, while the Decision Tree model reached 0.721 (95 CI: 0.704–0.738). In contrast, the Neural Network model showed reduced and more scattered performance ($R^2 = 0.671$; 95 CI: 0.695–0.724), consistent with its sensitivity to sparse or heterogeneous data, as previously observed. To examine whether broader training data could improve localized predictions under such conditions, we expanded the training dataset by combining the full-scale experiment (one-year experiment presented above) with the small-scale test set, then evaluated generalization performance on a held-out portion of the small dataset, without including any explicit spatial location labels. As shown in Fig. 12f–i, all models exhibited reduced accuracy compared to earlier evaluations. Nonetheless, Random Forest and XGBoost maintained relatively good performance of 0.802 (95 CI: 0.792–0.811) and 0.793 (95 CI: 0.791–0.804) respectively, while Decision Tree reached R^2 of 0.717 (95 CI: 0.7–0.734). The Neural Network model again underperformed substantially ($R^2 = 0.435$; 95% CI 0.425–0.449). These findings indicate that, even in the absence of spatial metadata, tree-based models can still produce moderately accurate local predictions when trained on diverse, large-scale data. Together, these results underscore the importance of combining temporal density and spatial diversity in building generalizable evaporation prediction models and highlight the robustness of tree-based approaches like XGBoost under real-world data limitations.

4. Discussion

In this study, we aimed to evaluate whether incorporating spatial–temporal environmental heterogeneity into machine learning models (ML) and the FPME improves prediction accuracy in machine learning (ML) models and evaporation estimation using the FAO-56 Penman–Monteith equation (FPME) under real greenhouse conditions. Specifically, we tested two main hypotheses: (1) providing high-resolution spatial–temporal environmental data to the FPME would improve its evaluation accuracy. (2) ML algorithms trained on the same high-resolution dataset would outperform the FPME by capturing nonlinear and hysteretic environmental dynamics. While the first hypothesis was not supported, the second was clearly validated – ML models showed significantly high accuracy across a wide range of conditions and datasets. These results highlight some limitations of the conventional models under heterogeneous and sub-daily conditions and point toward the value of data-driven approaches for capturing fine-scale environmental effects on evaporation.

4.1. Microclimatic challenges to the FPME model

The greenhouse environment exhibited pronounced spatial–temporal heterogeneity, as demonstrated by variations in temperature and relative humidity across both horizontal and vertical sensor positions (Fig. 2). These microclimatic differences, driven by structural elements such as shading from overhead beams, solar heating of walls, directional airflow and moist mattresses flow heterogenicity (Nebbali et al., 2012), created unique environmental conditions around each evaporative pan, this underscores the need for dense environmental sensing to achieve reliable spatial resolutions. To resolve this variability, we implemented inverse distance weighting (IDW: Eq. (1)), which enabled reconstruction of fine-scale environmental dynamics above each evaporative pan (Fig. 3). Despite this high spatial–temporal resolution, we found that as few asthree well-placed sensors were sufficient to produce accurate interpolation (Suppl Fig. 1; $p < 0.001$), enabling high-resolution environmental modeling without exhaustive sensor deployment. Despite the improved spatial–temporal resolution, we could not improve the accuracy of the FPME evaluation. Daily estimation deviated from observation up to ±22%, and sub-hourly (3-minute) prediction presented moderate correlation ($R^2 = 0.637$) (Fig. 6, Fig. 7 respectively). Furthermore, the distribution of daily FPME evaluation was narrower and more uniform than that of the actual lysimeter evaporative data (Fig. 6b.1–b.4). This smoothing effect was caused by averaging meteorological values and reducing the magnitude of the sub-daily effect (Batalha et al., 2018). Importantly, no clear correlation was observed between VPD – a primary theoretical driver of evaporation and transpiration demand – and measured midday evaporation rates across the bench (Fig. 4 a–h). This mismatch suggests that a single-point of VPD cannot explain spatial or sub-hour variations in evaporative flux under real conditions, even when measured at high temporal resolution, likely due to hysteretic impacts (see below).

We suggest that the reduced accuracy of the sub-daily FPME evaluation, despite the increased data density and high-precision lysimeter measurements, arises from systematic hysteretic effects, namely a temporal lag between environmental forcing and the evaporative response. As shown in Fig. 5, evaporation exhibited clear and direction-dependent diurnal hysteresis with respect to VPD. Such behavior is consistent with the thermal inertia of the water body, where accumulated energy delays the phase transition from liquid to vapor, generating hysteresis between environmental forcing and evaporation (Cui et al., 2016; McCUEN and Asmussen,

2009).

Importantly, these hysteresis patterns were not confined to a single day but recurred consistently across days with comparable diurnal environmental variables (See Suppl Fig. 4), indicating that the observed lag represents a stable system behavior rather than episodic variability.

We suggest that these non-linear dynamics cannot be captured by the linearized structure of the FPME. As others have shown (Leca et al., 2011; Paw U, 1992; Widmoser, 2009), the inability to account for temporal effects introduces substantial estimation error, particularly at finer time scales. Our findings further indicate that using daily or sub-daily averages to drive FPME computations reduces accuracy, as these averages smooth phase lags and nonlinear responses, thereby failing to capture the observed diurnal hysteresis in ET (Widmoser, 2009). Consequently, even with high-resolution environmental measurements, FPME remains limited in its capacity to capture real-time evaporation under heterogeneous and dynamically evolving conditions. Importantly, this hysteresis implies that model residuals are not random but are structured by time of day and microclimatic state, with the largest deviations occurring under sub-daily conditions characterized by strong phase lags between atmospheric forcing and evaporation. Together, these results highlight hysteresis as a mechanistic limitation of physically based empirical formulations and motivate the use of data-driven modeling frameworks, such as machine-learning approaches, that can explicitly learn non-linear relationships, temporal dependencies, and lagged system responses.

4.2. *Modelling evaporation under hysteretic and multifactorial conditions*

Models trained on the full dataset of interpolated environmental data, including both indirect ambient variables and direct measurements of water temperature, achieved consistently high prediction accuracy (Fig. 8a–d). The inclusion of water temperature as an input variable introduced additional variance that was not optimally exploited by the Neural Network architecture. Specifically, this effect is evident in the reduction of the R^2 value, from 0.955 when water temperature was excluded, down to 0.950 when it was included, demonstrating that the supposedly direct signal may introduce state-dependent variance that is not optimally exploited by all model architectures equally. Importantly, this does not imply that water temperature is redundant, but rather that its hysteretic, state-dependent nature may not be optimally captured by feed-forward neural architectures under fixed hyperparameter settings. This contrast highlights the challenge of incorporating such input into models whose architecture and sensitivity to input structure may limit their ability to exploit state-dependent inputs under fixed hyperparameter settings (Salahudin et al., 2023). To better understand the logic behind the model outputs and their links to our hypothesis, we analyzed feature importance rankings across all algorithms (Fig. 9). Although some consider these models as “black box” (Azodi et al., 2020), such ranking can reveal the relative contributions of each input. Interestingly, solar radiation emerged as the most influential feature in most models, except for XGBoost. This result is likely due to redundancy, as radiation was measured using a single sensor and assumed to be uniformly distributed, creating repeated identical values across all evaporative pans in each timestamp. This may have biased the model toward this feature.

Wind speed, in contrast, consistently ranked among the least informative features across all models. This result is coherent with the experimental design – the greenhouse had a controlled, low, and constant airflow. This microclimate in such a setting is decoupled from convective dynamics, making radiative and temperature factors more dominant – consistent with prior studies in greenhouse and screenhouse environments (Varga-Haszonits et al., 2022; Yohanani et al., 2022). Importantly, the hierarchy of features confirmed that short-term temporal variables – particularly hour of day and day of year – were consistently among the top three predictors of evaporation across all models (Fig. 9). In the most accurate model (XGBoost), hour ranked even higher than pan water temperature (Fig. 9, c, e). This dominance of temporal features suggests that the models internally capture hysteretic effects, where the timing of environmental forcing modulates evaporation through thermal memory. Although water temperature is theoretically the most direct physical driver of evaporation, its lower rankings (fourth) support the idea that hysteretic dynamics, in which changes in air temperature precede shifts in water temperature (Supplementary Fig. 4c), play a central role in determining real-time evaporative rates. This observation further supports the conclusion that evaporation is governed by temporally lagged processes rather than instantaneous environmental drivers alone. This apparent discrepancy can be explained by the fact that water temperature acts as an integrated state variable, whose predictive contribution partially overlaps with temporal encoding under high-frequency sampling. To further examine this relationship, we trained the models using a reduced feature set consisting only of water temperature and temporal variables. Under this minimal input configuration, the XGBoost model retained high predictive accuracy ($R^2 = 0.978$; Fig. 10 a–d), with hour of day emerging as the most informative variable (Fig. 10e). These findings validate that the delayed, time-Sstructure response of the evaporative pan is not only present but central to accurate prediction of dynamic evaporation under heterogeneous conditions. While temporal structure can partially substitute for direct thermal measurements, this does not negate the physical relevance of water temperature when the full environmental context is considered. Together, these results clarify that correlations between water temperature and ambient variables reflect physical coupling and hysteresis rather than redundancy, and that different machine-learning architectures vary in their ability to exploit such state-dependent information.

4.3. *Temporal limitations of the machine-learning algorithms*

High-resolution temporal sampling proved essential for improving prediction accuracy across all ML models. Moreover, when evaluating the effect of sampling frequency, we observed that higher sampling frequency significantly enhanced prediction performance (Fig. 11a). The accuracy of models declined when sampling intervals increased, reinforcing the importance of high-frequency data collection (Viggiano et al., 2021). These findings align with signal theory: to capture transient phenomena like evaporation, the sampling frequency should be at least twice the rate of the dominant fluctuation, consistent with Nyquist principles (Detwiler et al.,

2023). Yet, this high sampling frequency has its cost in terms of storage requirements and significant battery consumption. This trend is reflected by the logarithmic fit of Fig. 11a, which showed R^2 values of 0 0.835 for Decision Tree, 0.877 for Random Forest, 0.820 for XGBoost, and 0.418 for the Neural Network. This fit indicates that accuracy approaches a plateau at a sampling interval of approximately 27 min. The volume of training data also played a substantial role in model stability and accuracy (Fig. 11b). While tree-based models like random forest and XGBoost maintained high performance with as little as 10% of the full dataset ($\sim 3.10 * 10^5$ samples), the Neural Network exhibited high fluctuations, stabilized around ~80% of the dataset ($2.41 * 10^6$), this data dependency likely explains its decrease in accuracy under data-limited condition in other part of the experiment. This instability reflects a structural difference between the Neural Network and the tree-based models. The feed-forward Neural Network used here contains a relatively large number of trainable parameters, which become insufficiently constrained under sparse or temporally reduced training data. As a result, model performance becomes sensitive to noise and initialization. Hyperparameters and regularization settings were intentionally kept constant across all experiments to ensure reproducibility and comparability, but this constraint limits the Neural Network's ability to adapt to reduced-data regimes, in contrast to tree-based models that are inherently more robust under such conditions (Poulinakis et al., 2023). These results are consistent with previous work showing that Neural Networks often underperforms tree-based algorithms in small to medium dataset but may surpass them with sufficient data volume (Azodi et al., 2020; Kraus et al., 2020). To confirm that the models can achieve high accuracy with a substantially smaller dataset, we conducted an independent experiment using only three evaporative pans over a 30-day period, located in a separate section of the greenhouse. Despite the limited spatial coverage and shorter duration, prediction accuracy remained high – particularly for Random Forest and XGBoost, Importantly, the 95% confidence intervals (CI) of these reduced-data scenarios remained relatively narrow (Suppl Table 2) indicating that their performance reflected stable behavior under reduced-data confirming the robustness of these models even under small datasets, in contrast, the wider 95% CI observed in the Neural Network highlighted its sensitivity to training data volume (Fig. 12 a–d). Although there was no significant difference between the two leading models, the Neural Network's accuracy declined sharply, likely due to the reduced dataset size, and due to the architecture and regularization setting (Table 1) which were intentionally kept constant across experiments to ensure reproducibility. However, this constraint might limit the network's ability to adapt, resulting in this performance reduction.

This independent modified experiment confirms that accurate predictions can still be achieved using limited data, provided temporal resolution is high and spatial heterogeneity is captured. The fact that just three evaporative pans spread across the bench, sampled over one month, provided high-accuracy predictions underscores the critical role of both spatial placement and sampling strategy in data-driven environmental modeling.

4.4. Spatial limitations and generalization capacity of predictive models

When applying models trained on the full dataset (Fig. 8; black line) to a new spatial domain – specifically a second bench on the opposite side of the greenhouse – the prediction unexpectedly failed to achieve acceptable predictive accuracy relative to direct evaporation measurements (see Suppl. Fig. 2). This discrepancy suggests that spatially localized training may not generalize well across heterogeneous environments, despite the high volume and year-long data sets. This result defines a clear spatial generalization boundary: high predictive accuracy achieved within a single bench does not imply transferability across benches that differ in radiative exposure, airflow patterns, or thermal structure, even within the same greenhouse. We hypothesize that variation in sensor placement and pan position led to differing thermal responses, shaped by the greenhouse's uneven spatial structure and dynamic microclimatic gradients through the day. These findings indicate that the observed generalization failure arises from genuine microclimatic heterogeneity rather than from insufficient data volume or model complexity. To evaluate this, we re-trained the models using data from both benches, combining original and new spatial domains into unified datasets (Fig. 12f–j). While the accuracy decreases across all models – likely due to increased heterogeneity across spatial–temporal measurements – the relative accuracy performance and 95% CI (Suppl. Table 2) remained consistent. Random Forest and XGBoost continued to outperform other models, while the Neural Network model again showed the lowest accuracy and the widest 95% CI, consistent with the sensitivity the Neural Network, amplifying its inherent instability under heterogeneous conditions. These findings substantiate our claim regarding the significance of spatial sampling in environmental research and indicate that model generalization in heterogeneous environments depends not only on mean performance but also on prediction stability. In this context, model performance was more sensitive to spatial coverage than to total data volume, defining clear limits to cross-bench generalization. Consequently, greater emphasis should be placed on the strategic distribution of measurement points, rather than simply increasing the quantity of sensors deployed or the volume of data amassed. Practically, this implies that allocating sensors across distinct thermal and radiative zones is more effective for improving cross-bench model portability than dense sampling within a single bench. Even a limited number of strategically placed sensors spanning multiple benches can substantially enhance spatial generalization.

Our findings highlight the ability of machine learning models to capture sub-daily and sub-hourly evaporation dynamics in greenhouse environments, even when trained on relatively small but well-structured datasets. This accuracy was achieved not by the sheer volume of data, but by combining high-frequency temporal sampling with strategically positioned sensors that reflect spatial heterogeneity. Such an approach underscores the value of fine-scale environmental monitoring in predictive modeling. From a design perspective, these results emphasize that spatial representativeness, rather than dataset size alone, should guide experimental layout and sensor deployment in data-driven greenhouse studies. Looking ahead, these methods may be extended to estimate more complex physiological processes – such as plant transpiration, crop coefficients, and yield – which are influenced by both environmental and biological regulation. While these tasks will likely introduce new challenges, the growing accessibility of low-cost, wireless

microsensors opens the door to dense, real-time monitoring of plant-environment interaction. Together, these developments hold promises for building robust, scalable predictive frameworks in modern, data-driven agriculture.

4.5. Limitations, recommendations, and practical implications

This study was conducted in a semi-controlled Mediterranean greenhouse that regulated airflow; therefore, extrapolation to open-field or contrasting climatic conditions require adaptation and validation. Model transferability was spatially constrained, as cross-bench generalization declined due to micro-scale thermal gradients, and light intensity was measured at a single location and assumed spatially uniform, potentially underrepresenting spatial heterogeneity. Neural-network performance was sensitive to reduced data volume under fixed hyperparameters. Future work should validate the framework across contrasting climatic regions and evaluate feature importance stability under varying environmental conditions, and deploy denser microclimatic networks, including low-cost IoT-based sensors. Extending the framework to daily and sub-daily whole-plant transpiration will require precise integration of physiological measurements capturing stomatal and hydraulic regulation. Practically, the results indicate that predictive robustness depends more on high-frequency temporal sampling and strategic spatial placement than on data size alone. Distributed, high-resolution environmental monitoring provides a stronger basis for data- driven irrigation scheduling than single point measurements.

5. Conclusion

Over extended temporal scales, predictive performance is constrained primarily by temporal resolution, spatial representativeness, and hysteretic system dynamics, rather than by data volume alone. This reflects the fact that evaporation under greenhouse conditions is governed by time-dependent, hysteretic dynamics rather than by instantaneous atmospheric demand. Improving long-term accuracy requires high-frequency sampling, spatially distributed sensor placement across heterogeneous microclimatic zones, and modeling frameworks capable of capturing non-linear and lagged responses. Our methods and results contribute to the United Nations Sustainable Development Goals (SDGs) (SDG's) by their potential to improve irrigation efficiency through accurate micro-meteorological sensing, encouraging more efficient use of technological and water resources (SDG 12: Responsible Consumption and Production) and strengthening climate-resilient agricultural management (SDG 13: Climate Action). Together, these considerations define practical design guidelines for a scalable, data-drive framework for evaporation prediction that moves beyond VPD.

CRediT authorship contribution statement

Menachem Moshelion: Writing – review & editing, Supervision. **Averbuch Nir:** Writing – review & editing, Writing – original draft, Visualization, Software, Formal analysis.

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

The work was funded by the Israel Ministry of Innovation, Science and Technology (Grant nos. 0002122 and 001897).

This project was funded by the Planning and Budgeting Committee of the Council for Higher Education in Israel (VATAT) as part of the program "Israeli Center for Digital Agriculture.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.ejrh.2026.103324.

Data availability

Data will be made available on request.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X., Research, G., 2016. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems.

Allen, R.G., Pereira, L.S., 1998. Crop Evapotranspiration. Guidelines for Computing Crop Water Requirements.

Azodi, C.B., Tang, J., Shiu, S.H., 2020. Opening the black box: interpretable machine learning for geneticists. Trends Genet. 36, 442–455. https://doi.org/10.1016/J.TIG.2020.03.005.
Balasubramanian, H.K., Thirugnanam, H., 2023. Neural networking to predict sap flow using AI-synthesized relative meteorological data. In: ju2023 3rd International Conference on Intelligent Technologies (CONIT). IEEE, pp. 1–7. https://doi.org/10.1109/CONIT59222.2023.10205886.
Başağaoğlu, H., Chakraborty, D., Winterle, J., 2021. Reliable evapotranspiration predictions with a probabilistic machine learning framework. Water 13, 557. https://doi.org/10.3390/w13040557.
Batalha, M.S., Barbosa, M.C., Faybishenko, B., Van Genuchten, M.T., 2018. Effect of temporal averaging of meteorological data on predictions of groundwater recharge. Journal of Hydrology and Hydromechanics 66 (2), 143–152. https://doi.org/10.1515/johh-2017-0051.
Cabrera-Bosquet, L., Fournier, C., Brichet, N., Welcker, C., Suard, B., Tardieu, F., 2016. High-throughput estimation of incident light, light interception and radiation-use efficiency of thousands of plants in a phenotyping platform. New Phytol. 212, 269–281. https://doi.org/10.1111/NPH.14027;WGROUP:STRING:PUBLICATION.
Chen, T., Guestrin, C., 2016. XGBoost: a scalable tree boosting system. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 13–17-August-2016, pp. 785–794. https://doi.org/10.1145/2939672.2939785/SUPPL_FILE/KDD2016_CHEN_BOOSTING_SYSTEM_01-ACM.MP4.
Chen, Z., Sun, S., Zhu, Z., Jiang, H., Zhang, X., 2019. Assessing the effects of plant density and plastic film mulch on maize evaporation and transpiration using dual crop coefficient approach. Agric. Water Manag. 225, 105765. https://doi.org/10.1016/J.AGWAT.2019.105765.
Cui, Y., Liu, Y., Gan, G., Wang, R., 2016. Hysteresis Behavior of Surface Water Fluxes in a Hydrologic Transition of an Ephemeral Lake. https://doi.org/10.1029/2019JD032364.
Das, A., 2022. Multivariate statistical approach for the assessment of water quality of Mahanadi basin, Odisha. Mater. Today Proc. 65, A1–A11. https://doi.org/10.1016/j.matpr.2022.08.146.
Das, A., 2024. An innovative approach for quality assessment and its contamination on surface water for drinking purpose in Mahanadi River Basin, Odisha of India, with the integration of BA-WQI, AHP-TOPSIS, FL-DWQI, MOORA, and RF methodology. Appl. Water Sci. 14 (12), 263. https://doi.org/10.1007/s13201-024-02326-9 (-.
Das, D.A., 2025. Harnessing hydro chemical characterization of surface water using water quality indices and machine learning – driven water quality modelling with special emphasis on side – stream pollution. Desalin. Water Treat. 324, 101592. https://doi.org/10.1016/j.dwt.2025.101592.
Das, A., 2025. Surface water quality evaluation impacting drinking water sources and sanitation using water quality index, multivariate techniques, and interpretable machine learning models in Mahanadi River, Odisha (India). Environ. Geochem. Health 47. https://doi.org/10.1007/s10653-025-02806-0.
de Meneses, Klara Cunha, Aparecido, L.E.D.O., de Meneses, Kamila, Cunha, de Farias, M.F., 2020. Estimating potential evapotranspiration in maranhão state using artificial neural networks. Rev. Bras. de Meteorol. 35, 675–682. https://doi.org/10.1590/0102-77863540072.
Detwiler, S., Barbosa, E., Steudel, K., Tippmann, J., Ward, C., West, B., 2023. Multi-sensor collaborative sampling schemes to reconstruct undersampled mechanical system signals for machinery fault detection. Conf. Proc. Soc. Exp. Mech. Ser. 51–60. https://doi.org/10.1007/978-3-031-04098-6_6/FIGURES/6.
Es-Sabery, F., Hair, A., Qadir, J., Sainz-De-Abajo, B., Garcia-Zapirain, B., Torre-DIez, I., 2021. Sentence-level classification using parallel fuzzy deep learning classifier. IEEE Access 9, 17943–17985. https://doi.org/10.1109/ACCESS.2021.3053917.
Ferreira, L.B., da Cunha, F.F., de Oliveira, R.A., Fernandes Filho, E.I., 2019. Estimation of reference evapotranspiration in Brazil with limited meteorological data using ANN and SVM – a new approach. J. Hydrol. 572, 556–570. https://doi.org/10.1016/J.JHYDROL.2019.03.028.
Ghiat, I., Mackey, H.R., Al-Ansari, T., 2021. A review of evapotranspiration measurement models, techniques and methods for open and closed agricultural field applications. Water 13, 2523. https://doi.org/10.3390/W13182523.
Kim, S.J., Bae, S.J., Jang, M.W., 2022. Linear regression machine learning algorithms for estimating reference evapotranspiration using limited climate data. Sustainability 14, 11674. https://doi.org/10.3390/SU141811674.
Kraus, M., Feuerriegel, S., Oztekin, A., 2020. Deep learning in business analytics and operations research: models, applications and managerial implications. Eur. J. Oper. Res. 281, 628–641. https://doi.org/10.1016/J.EJOR.2019.09.018.
Leca, A., Parisi, L., Lacointe, A., Saudreau, M., 2011. Comparison of Penman–Monteith and non-linear energy balance approaches for estimating leaf wetness duration and apple scab infection. Agric. For. Meteor. 151, 1158–1162. https://doi.org/10.1016/J.AGRFORMET.2011.04.010.
Li, J., Cheng, K., Wang, S., Morstatter, F., Trevino, R.P., Tang, J., Liu, H., 2017. Feature selection: a data perspective. ACM Comput. Surv. 50. https://doi.org/10.1145/3136625.
Liao, L., Zhang, L., Bengtsson, L., 2008. Soil moisture variation and water consumption of spring wheat and their effects on crop yield under drip irrigation. Irrig. Drain. Syst. 22, 253–270. https://doi.org/10.1007/S10795-008-9055-5.
Markulj Kulundži, A., Viljevac Vuleti, M., Matoša Kočar, M., Miji, A., Varga, I., Sudari, A., Cesar, V., Lepeduš, H., Van Ha, C., Golam Mostofa, M., Saha, G., Roy Choudhury, S., 2021. The Combination of Increased Temperatures and High Irradiation Causes Changes in Photosynthetic Efficiency. https://doi.org/10.3390/plants10102076.
McCUEN, R.H., Asmussen, L.E., 2009. Estimating the effect of heat storage on evaporation rates. Hydrol. Sci. J. 18, 191–196. https://doi.org/10.1080/02626667309494026.
Nebbali, R., Roy, J.C., Boulard, T., 2012. Dynamic simulation of the distributed radiative and convective climate within a cropped greenhouse. Renew. Energy 43, 111–129. https://doi.org/10.1016/J.RENENE.2011.12.003.
Paw U, K.T.P.U., 1992. A discussion of the Penman form equations and comparisons of some equations to estimate latent energy flux density. Agric. For. Meteor. 57, 297–304. https://doi.org/10.1016/0168-1923(92)90125-N.
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É., 2012. Scikit-learn: machine learning in python. J. Mach. Learn. Res. 12, 2825–2830.
Poulinakis, K., Drikakis, D., Kokkinakis, I.W., Spottswood, S.M., 2023. Machine-learning methods on noisy and sparse data. Mathematics 11, 236. https://doi.org/10.3390/math11010236.
Salahudin, H., Shoaib, M., Albano, R., Inam Baig, M.A., Hammad, M., Raza, A., Akhtar, A., Ali, M.U., 2023. Using ensembles of machine learning techniques to predict reference evapotranspiration (ET0) using limited meteorological data. Hydrology 10, 169. https://doi.org/10.3390/HYDROLOGY10080169.
Schlesinger, W.H., Jasechko, S., 2014. Transpiration in the global water cycle. Agric. For. Meteor. 189–190, 115–117. https://doi.org/10.1016/J.AGRFORMET.2014.01.011.
Shamshiri, R.R., Jones, J.W., Thorp, K.R., Ahmad, D., Man, H.C., Taheri, S., 2018. Review of optimum temperature, humidity, and vapour pressure deficit for microclimate evaluation and control in greenhouse cultivation of tomato: a review. Int. Agrophys. 32, 287–302. https://doi.org/10.1515/intag-2017-0005.
Van Dijk, M., Morley, T., Rau, M.L., Saghai, Y., 2021. A meta-analysis of projected global food demand and population at risk of hunger for the period 2010–2050. Nat. Food. https://doi.org/10.1038/s43016-021-00322-9.
Varga-Haszonits, Z., Szalka, É., Szakál, T., Varga-Haszonits, Z., Szalka, É., Szakál, T., 2022. Determination of reference evapotranspiration using Penman-Monteith method in case of missing wind speed data under subhumid climatic condition in Hungary. Atmos. Clim. Sci. 12, 235–245. https://doi.org/10.4236/ACS.2022.122014.
Viggiano, M., Geraldi, E., Cimini, D., Di Paola, F., Gallucci, D., Gentile, S., Larosa, S., Nilo, S.T., Ricciardelli, E., Romano, F., 2021. The role of temporal resolution of meteorological inputs from reanalysis data in estimating air humidity for modelling applications. Agric. For. Meteor. 311, 108672. https://doi.org/10.1016/j.agrformet.2021.108672.
Wei, Z., Yoshimura, K., Wang, L., Miralles, D.G., Jasechko, S., Lee, X., 2017. Revisiting the contribution of transpiration to global terrestrial evapotranspiration. Geophys. Res. Lett. 44, 2792–2801. https://doi.org/10.1002/2016GL072235;PAGEGROUP:STRING:PUBLICATION.
Widmoser, P., 2009. A discussion on and alternative to the Penman-Monteith equation. Agric. Water Manag. 96, 711–721. https://doi.org/10.1016/J.AGWAT.2008.10.003.

Yaaran, A., Negin, B., Moshelion, M., 2019. Role of guard-cell ABA in determining steady-state stomatal aperture and prompt vapor-pressure-deficit response. Plant Sci. 281, 31–40. https://doi.org/10.1016/J.PLANTSCI.2018.12.027.

Yao, M., Gao, M., Wang, J., Li, B., Mao, L., Zhao, M., Xu, Z., Niu, H., Wang, T., Sun, L., Niu, D., 2023. Estimating evapotranspiration of greenhouse tomato under different irrigation levels using a modified dual crop coefficient model in Northeast China. Agriculture 13, 1741. https://doi.org/10.3390/AGRICULTURE13091741.

Yohanani, E., Frisch, A., Lukyanov, V., Cohen, S., Teitel, M., Tanny, J., 2022. Estimating evapotranspiration of screenhouse banana plantations using artificial neural network and multiple linear regression models. Water 14, 1130. https://doi.org/10.3390/W14071130.