remote sensing

MDPI

Article

Physically Constrained Estimation of Dynamic Crop Coefficients Using Multisensor Satellite and Field Telemetry Data

Umer Tanveer [1], Hashim Ali [1], Jamal Ahmed [1], Abdu Salam [1], Farhan Amin [2,*], Isabel de la Torre [3], Lázaro Javier Hernández Rodríguez [4] and Pablo Herrero García [4]

[1] Department of Computer Science, Abdul Wali Khan University, Mardan 23200, Pakistan; umertanveer@awkum.edu.pk (U.T.); hashimali@awkum.edu.pk (H.A.); jamal.ahmed@awkum.edu.pk (J.A.); abdusalam@awkum.edu.pk (A.S.)
[2] School of Computer Science and Engineering, Yeungnam University, Gyeongsan 38541, Republic of Korea
[3] eHealth and Telemedicine Group, University of Valladolid, 47011 Valladolid, Spain; isator@uva.es
[4] Department of Projects, Universidad Europea del Atlántico, Isabel Torres 21, 39011 Santander, Spain; lazaro.hernandez@uneatlantico.es (L.J.H.R.); pablo.herrero@uneatlantico.es (P.H.G.)
* Correspondence: farhanamin10@hotmail.com

Highlights

What are the main findings?

- The proposed AquaVolt-AI Physics-Informed Machine Learning (PIML) framework eliminated crop-coefficient-bound violations (0% vs. 14% for the unconstrained baseline) and reduced RMSE by 14.1% (0.3263 vs. 0.3801) in short-term temporal validation.
- Accounting for spatial autocorrelation (ICC = 0.65) reduced the nominal test sample size from N = 200 to an estimated effective sample size of 12.0. The accuracy difference between AquaVolt-AI and the unconstrained baseline was not statistically significant after design-effect correction ($p = 0.5492$).

What are the implications of the main findings?

- The framework's hard constraints on neural-network corrections (±0.15) and final crop-coefficient predictions (0.15–1.20) ensure compliance with predefined agronomic bounds, potentially improving the reliability of satellite-assisted irrigation decision-support systems.
- We conducted experiments, and validation using repeated observations from the same agricultural fields demonstrates short-term temporal transfer rather than spatial generalization. Larger independent datasets and field-blocked validation are needed to establish geographic transferability and reliably assess predictive performance.

Abstract

Accurate estimation of crop evapotranspiration (ETc) plays an important role in sustainable agricultural water management, precision irrigation planning, and food security under a dynamically changing climate. Physics-based models such as the FAO-56 Penman–Monteith equation have a sound theoretical background but are limited in modeling real-time dynamics in vegetation, owing to their use of static crop coefficient (Kc) tables. Meanwhile, ML models trained between the satellite-derived vegetation indices and crop coefficients are prone to physical bound violations and poor spatial generalizability. In this study, the Physics-Informed Machine Learning (PIML) framework is introduced, which consists of connecting the data from Sentinel-2 multispectral reflectance with the data from MODIS gap-filled time-series and the thermal-based reference evapotranspiration by NASA ECOSTRESS within a hard-bounded residual structure, $Kc = \text{clip}(Kc, \text{prior} + \text{clip}(f\theta(x), -0.15, 0.15), 0.15, 1.20)$, where the neural network's correction term resides between FAO-56 agronomic bounds. Although the held-out validation set ($N = 200$) was intentionally kept small due to

Check for updates

Academic Editor: Guido D'Urso

Received: 16 September 2026
Revised: 5 November 2026
Accepted: 12 November 2026
Published: 14 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

the limited number of cloud-free ECOSTRESS overpasses, an operational telemetry system recorded hourly data from four crop fields at the Sustainable Agriculture Facility (RSRF) in California. Because the same four fields and 25 sectors are visited across both training and testing periods, the experimental evaluation represents an intra-site, short-term temporal forward-transfer test rather than a true spatial generalization. The limited sample size (five cloud-free dates) precluded true spatial generalization methods such as leave-one-field-out or field-blocked cross-validation, so we kept all four crops in the training set. Satellite data for the same field are highly spatially correlated; therefore, we used the ICC (intraclass correlation coefficient) to adjust for non-independence, reducing N = 200 to n_{eff} =12.0. Three models were benchmarked: an empirical baseline with an RMSE = 0.3801, a constrained PIML network (RMSE = 0.3263), and a global training-mean predictor (RMSE = 0.3257). Model B removed all bound violations in the physical domain (compared to 14% for Model A), reduced RMSE by 14.1% from the baseline, and produced results similar to the global mean; the design-effect-corrected *p*-value (0.5492) confirms the pilot comparison is statistically underpowered. When validating remote sensing applications with small, clustered samples, these results show that hard physical constraints eliminate physically implausible predictions during temporal forward projection, while highlighting the need to correct spatial autocorrelation and requiring dedicated field-blocked evaluations for future claims of spatial generalization.

Keywords: crop evapotranspiration; land; machine learning; agricultural; vegetation

1. Introduction

Accelerating climate change, rapid population growth, and catastrophic depletion of subsurface aquifer systems [1–3] pose unique challenges for agricultural water management worldwide. Since the dawn of the planetary boundaries framework, global freshwater boundaries have been breached [1], making machine learning solutions for climate change mitigation and adaptation [2] and irrigation-efficiency interventions a critical global priority. The carbon and water cycles are closely coupled: atmospheric carbon dioxide levels rise, regional temperature anomalies strengthen, droughts become more frequent, and agricultural growing seasons become more unpredictable [4–6]. Agriculture uses about 70% of the world's freshwater [3]. In fast-growing regions such as South Asia, South-East Asia, and Sub-Saharan Africa, groundwater depletion rates have increased significantly over the last 30 years, threatening the long-term sustainability of irrigation-based agricultural economies. In this context, measuring and predicting crop water needs precisely has shifted from an operational task to a policy mandate.

Estimating evapotranspiration (ET) has evolved over a century of scientific development, from empirical approximations to physically rigorous multisensor satellite frameworks. The formulations used only temperature and solar radiation as inputs and thus did not include aerodynamic and humidity fluxes in evaporative demand. However, these temperature-based models are still widely used in data-limited regions but consistently fail in arid and semi-arid regions, where advective wind energy plays a large role in atmospheric water demand.Theradiation-based formulation improved on this by including net radiation and soil heat flux, but it still did not account for aerodynamic resistance to vapor transport.

The current international standard for reference evapotranspiration estimation is the FAO-56 Penman–Monteith equation [3,6], a physically complete energy balance equation that includes net radiation, soil heat flux, vapor pressure deficit, aerodynamic resistance, and canopy resistance. The FAO-56 framework also integrates the crop coefficient, Kc: it divides the problem into two parts, meteorological evaporative demand (reference evap-

otranspiration, ET0, for a standard grass surface) and crop biological aspects (tabular Kc values from decades of lysimeter research). Although simple, this dual-component framework can introduce significant errors in practice when crop variety, planting density, microclimatic conditions, and soil type differ from the experimental conditions.

To overcome the spatial and temporal constraints of lysimeter measurements, researchers have used remote sensing extensively to monitor canopy dynamics and develop Kc tables [7–9]. Starting with the Landsat satellites in 1972 and later the ESA Sentinel-2 constellation in 2015, vegetation has been monitored globally at 10–30 m resolution in a consistent manner. NDVI from the red and near-infrared reflectance channels correlates well with crop canopy development, fractional ground cover, and leaf area index, making it a good proxy for dynamically estimating Kc [9]. Furthermore, recent advances in hyperspectral imaging technology provide continuous spectral responses that capture detailed crop morphological and physiological variations for agricultural monitoring [10].

Meanwhile, unmanned aerial vehicles (UAVs) with hyperspectral sensors have proven useful for monitoring crops at the sub-field level with centimeter-scale resolution, enabling identification of areas of water stress and irrigation deficiencies [11,12].

Over the last two decades, satellite-scale energy-balance formulations for ET retrieval have improved greatly. Methods such as METRIC (Mapping Evapotranspiration at High Resolution with Internalized Calibration), SEBAL (Surface Energy Balance Algorithm for Land), and TSEB (Two-Source Energy Balance) utilize the land surface temperature (LST) derived from a thermal infrared sensor to separate the net radiation into the latent heat, sensible heat, and soil heat fluxes [13]. These methods have physical requirements but are not easily automated at continental scales because they require intensive, site-specific calibration of cold and hot pixel anchors. In addition, cloud contamination is a unique problem for thermal satellite observations, creating substantial data gaps during the most important agricultural growing periods.

The Internet of Things (IoT) revolution has transformed agricultural monitoring, enabling densely distributed, low-cost environmental sensors across field networks [14–17]. Smart telemetry grids can now measure ambient air temperature, relative humidity, wind speed, solar irradiance, soil moisture, and electrical conductivity at sub-hourly intervals from hundreds of sensor nodes simultaneously. These networks, coupled with satellite data streams through cloud computing platforms, produce vast operational databases of spatiotemporal variability in agricultural microenvironments. However, large operational databases do not imply large databases of ground-truth real evapotranspiration measurements to validate models; only specialized equipment, such as a weighing lysimeter or an eddy covariance tower, provides such direct physical measurements. Satellite thermal overpasses (e.g., ECOSTRE) offer a more spatially extensive, though less direct, alternative: rather than measuring evapotranspiration directly, they provide a physically based model estimate that this study uses as a reference when ground-based flux measurements are unavailable.

AI and ML can effectively combine different data streams to support agricultural irrigation scheduling and water management [18–20]. Neural networks, gradient-boosted trees, and ensemble methods outperform the traditional empirical models when trained on sufficient historical data [21]. One fundamental problem with traditional data-driven ML models, however, is that they can produce physically unrealistic answers when applied to data outside their training range or to atypical test values. An unconstrained neural network can predict values such as Kc = 0.05 for dense corn during peak summer and Kc = 1.8 for a bare soil field after harvest, which are beyond reasonable agronomic practices. If automated irrigation scheduling systems compound factors that increase prediction error (PE), they can cause significant crop stress or over-irrigation, offsetting irrigation efficiency benefits.

Physics-Informed Machine Learning (PIML) resolves this conflict by integrating domain-specific physics into the learning objective, model design, or model output [13,22,23]. A growing number of hydrologists and water resources engineers have recently realized that the performance of PIML models is superior in hydrologic regimes where very few or no data are available when compared with that of purely empirical models [23–28]. PIML inherently narrows the model's hypothesis space, meaning that, in addition to preventing overfitting, the physically viable hypothesis set transfers better across locations and climate zones [29].

Soft physical constraints in existing PIML applications to crop-coefficient estimation—typically a penalty term in the loss function—do not necessarily prevent physically implausible outputs, and none of these approaches have combined this with a validation process that accounts for the strong spatial autocorrelation of pixel-level satellite data. This raises an open question: whether a hard, architecturally enforced bound, combined with multisensory fusion (Sentinel-2, MODIS, and ECOSTRESS) and a spatially rigorous accuracy assessment, can eliminate physically impossible predictions without compromising the honesty of the reported statistical significance.

The following are the main contributions of this paper:

Unlike naive post hoc truncation (evaluated in our ablation suite as Model A_{clip}), which artificially clamps out-of-bounds predictions without correcting internal network representations, AquaVolt-AI integrates the physical boundary directly into the learning objective. The quadratic penalty in Equation (13) provides nonzero corrective gradients exclusively when the pre-clip intermediate prediction $\tilde{K}_c$ transgresses the admissible envelope $[0.15, 1.20]$, while remaining inactive inside the feasible zone. Consequently, the network learns to adapt its internal weights to generate residual corrections that remain structurally feasible, rather than relying solely on post hoc truncation after inference.

This study introduces AquaVolt-AI, a physically constrained residual-learning framework that combines an FAO-56 crop-coefficient prior with a bounded neural-network correction to estimate dynamic crop coefficients and crop evapotranspiration. The correction is limited to $(-0.15 \leq \Delta K_c \leq 0.15)$, while the final prediction is restricted to the agronomically plausible range $(0.15 \leq \hat{K}_c \leq 1.20)$, thereby preventing unrealistic outputs. The framework integrates Sentinel-2 multispectral reflectance, MODIS-based temporal gap filling, meteorological data, hourly field telemetry, and ECOSTRESS-derived thermal observations in a unified multisensor workflow. Beyond predictive accuracy, the study evaluates the physical consistency of model outputs and shows that AquaVolt-AI eliminates the crop-coefficient bound violations produced by the empirical baseline. The study also introduces a spatial-autocorrelation-aware validation procedure based on the intraclass correlation coefficient and design effect, reducing the nominal validation sample size from $(N = 200)$ to an effective sample size of $(n_{\text{eff}} = 12.0)$. Crucially, we make explicit the distinction between the short-term temporal forward transfer evaluated here (across identical fields and sectors over consecutive days) and true spatial generalization. While spatial generalization claims in remote sensing strictly require field-blocked or leave-one-field-out (LOFO) partitions across geographically detached sites, our constrained sample size of five overpasses necessitated an intra-site temporal design to retain all four crop types during training. Finally, it provides a transparent assessment of uncertainty by showing that, although AquaVolt-AI reduced RMSE relative to the empirical baseline, its performance was comparable to the global training-mean predictor and the corrected difference was not statistically significant. These contributions support more physically reliable machine-learning applications for evapotranspiration estimation while emphasizing the boundary between temporal transfer and spatial generalization in clustered satellite datasets.

2. Materials and Methods

2.1. Study Site Description

This study was primarily validated at the University of California, Davis (UCD) Russell Ranch Sustainable Agriculture Facility (latitude: 38.54°N; longitude: 121.87°W; elevation: 17 m above sea level (ASL)) (Figure 1). Russell Ranch is a 150-ha long-term agricultural research facility that was established in 1992 to assess sustainable farming systems under controlled management protocols. The facility includes four major crop management areas: organic corn, conventional corn, an alfalfa/legume rotation, and wheat production fields. These fields use drip and furrow irrigation, with automated soil moisture monitoring throughout the growing season. Such contrasting crop canopies and irrigation regimes create complex microclimatic gradients that govern the spatiotemporal dynamics of actual evapotranspiration across managed agricultural landscapes [30].

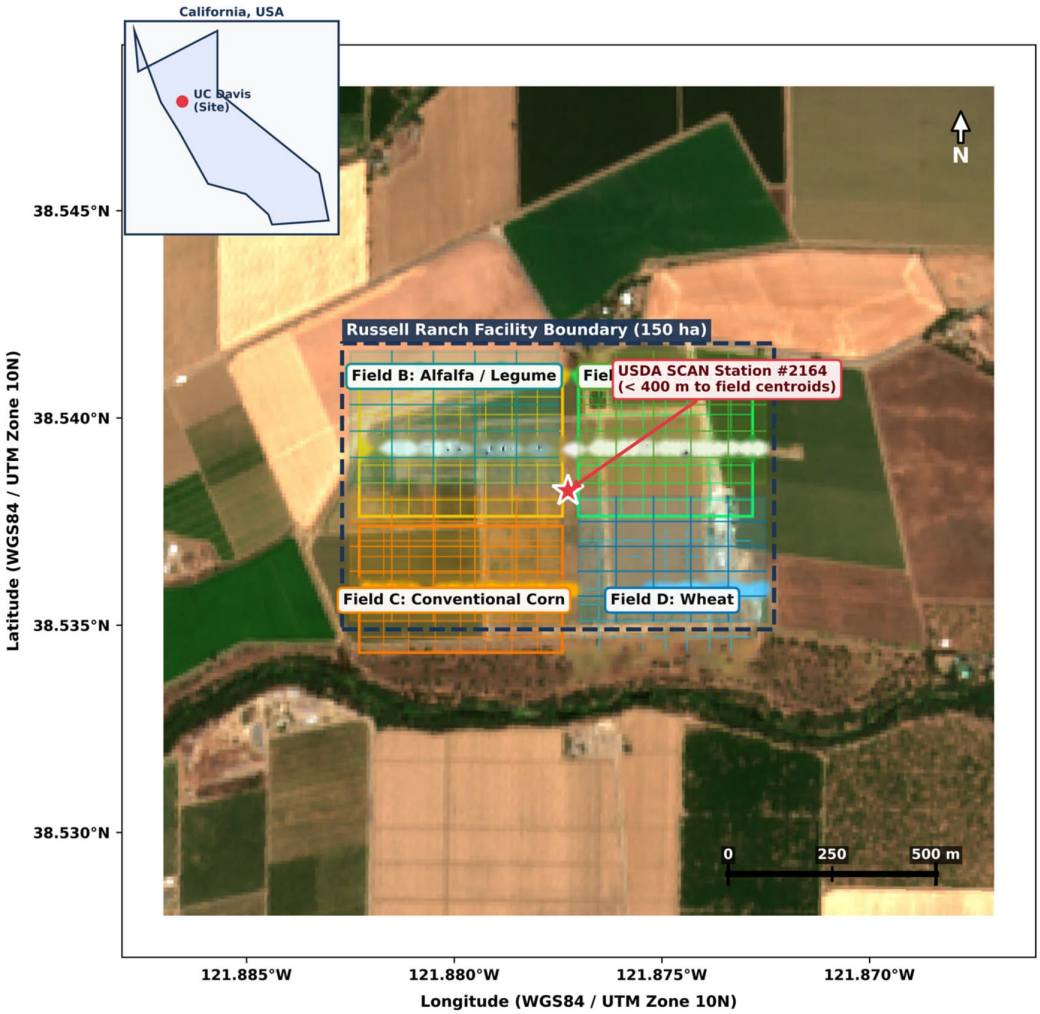

Figure 1. Spatial layout of the target agricultural fields at the Russell Ranch Sustainable Agriculture Facility, UC Davis, California. Field polygons are color-coded by crop type. The coordinate system is WGS84/UTM Zone 10N.

2.2. Multisensor Data Acquisition and Fusion Pipeline

Figure 2 shows the Schematic of the AquaVolt-AI PIML architecture. The AquaVolt-AI data fusion pipeline integrates four complementary observation streams: Sentinel-2 optical reflectance, NASA ECOSTRESS thermal evapotranspiration, ERA5-Land surface meteorology, and field-scale telemetry. Such integrated multi-source telemetry frameworks

are increasingly vital for deploying artificial intelligence in precision agriculture to optimize dynamic irrigation scheduling and crop water allocation [31].

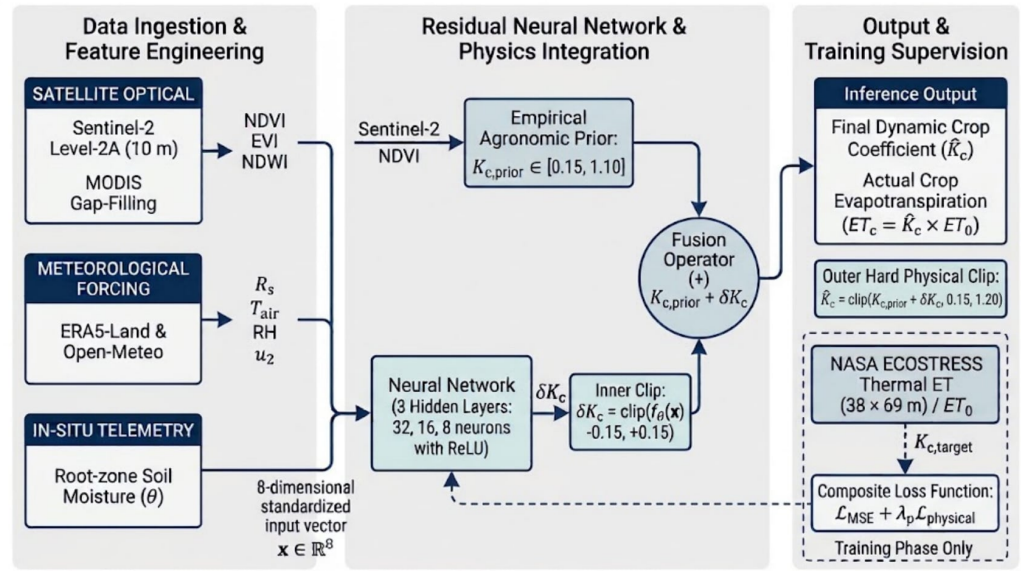

Figure 2. Schematic of the AquaVolt-AI PIML architecture. Satellite, weather, and soil features are processed by a residual MLP to generate $\delta K_c \in [-0.15, +0.15]$, which is fused with the FAO-56 prior into $\tilde{K}_c = K_{c,\text{prior}} + \delta K_c$. During training, $\tilde{K}_c$ is evaluated by the boundary penalty $L_$physical and passed through outer hard clipping to produce the final crop coefficient $\hat{K}_c \in [0.15, 1.20]$. NASA ECOSTRESS observations serve strictly as training targets, not inference input.

2.2.1. Sentinel-2 Multispectral Imagery

The primary vegetation spectral data source was ESA Sentinel-2 Level-2A (L2A) surface reflectance products. Sentinel-2 provides 13 spectral bands spanning the visible, near-infrared (NIR), and shortwave infrared (SWIR) spectrum at 10 m, 20 m, and 60 m spatial resolutions. High-resolution multispectral vegetation indices serve as primary optical predictors for machine learning models of crop canopy biophysical properties and management zones [32]. The NDVI was computed from the 10 m resolution Red (B04, 665 nm) and NIR (B08, 842 nm) bands:

$$NDVI = \frac{B08 - B04}{B08 + B04} \quad (1)$$

Atmospheric correction was applied via the Sen2Cor processor (version 2.11), which removes Rayleigh scattering, aerosol optical depth effects, and water vapor absorption from the top-of-atmosphere (TOA) radiance to produce surface reflectance. Cloud and cloud-shadow masking was performed using the Scene Classification Layer (SCL), which classifies each pixel into 11 land cover categories, including cloud medium probability, cloud high probability, cloud shadow, and water. Pixels classified as clouds or cloud shadows were excluded from all analyses. Remaining valid observations were resampled to a uniform 10 m grid using bilinear interpolation.

In addition to NDVI, we computed the Normalized Difference Water Index (NDWI) from green (B03) and NIR (B08) bands and the Red Edge Chlorophyll Index (ClRE) from the red-edge band (B05) to provide complementary information on canopy water content and chlorophyll concentration, respectively. The Enhanced Vegetation Index (EVI) was also computed using the following:

$$EVI = 2.5 \times \frac{B08 - B04}{B08 + 6B04 - 7.5B02 + 1} \quad (2)$$

Here, B02 is the blue band (490 nm). EVI is less susceptible to atmospheric residuals and canopy background effects than NDVI, particularly in high-biomass regions where NDVI saturates.

2.2.2. Meteorological Forcing from Open-Meteo and ERA5-Land

Meteorological forcing for the standardized FAO-56 ET_0 calculation was obtained from ERA5-Land reanalysis [33], accessed via the Open-Meteo API, at hourly temporal resolution and approximately 9 km spatial resolution. Five primary surface variables were extracted at the grid cell nearest the Russell Ranch facility (38.54° N, 121.87° W): (1) 2-m air temperature (°C), aggregated to daily mean (T_{mean}), maximum (T_{max}), and minimum (T_{min}); (2) 2-m dewpoint temperature (°C), used to derive actual vapor pressure (e_a, kPa); (3) downward surface solar radiation (shortwave radiation, R_s, hourly average in W m^{-2}), integrated over the 24-h diurnal cycle to daily total irradiance (MJ m^{-2} day^{-1}); (4) 10-m horizontal wind speed (m s^{-1}), converted to the 2-m reference height (u_2) using the FAO-56 logarithmic wind profile; and (5) surface atmospheric pressure (P, kPa), used to compute the psychrometric constant. Daily aggregates were computed using local Pacific Time (UTC−7) day boundaries.

Native reanalysis net radiation was not ingested directly. Instead, downward shortwave radiation (R_s) was used to compute reference net radiation (R_n) for the standardized FAO-56 grass reference surface, with fixed albedo $\alpha = 0.23$, as detailed in Section 2.5. The net radiation at the hypothetical grass reference surface is expressed in MJ m^{-2} day^{-1}. Rather than extracting unstandardized net radiation directly from the reanalysis grid, which reflects heterogeneous land cover and variable surface albedos, R_n was computed following the standardized FAO-56 formulation $R_n = R_{ns} - R_{nl}$. Here, R_{ns} is net solar (shortwave) radiation (MJ m^{-2} day^{-1}), calculated as: $R_{ns} = (1 - \alpha)R_s$. The fixed albedo $\alpha = 0.23$ represents the hypothetical green grass reference surface, and R_s is daily total downward solar radiation (MJ m^{-2} day^{-1}) derived from ERA5-Land. Net outgoing longwave radiation (R_{nl}, MJ m^{-2} day^{-1}) was computed using the FAO-56 Stefan–Boltzmann expression adjusted for vapor pressure and cloudiness: $R_{nl} = \sigma \left[\left(\left(T_{max,K}^4 + T_{min,K}^4\right)/2\right)\right] \left(0.34 - 0.14\sqrt{e_a}\right) \left[1.35(R_s/R_{so}) - 0.35\right]$. Here, $\sigma = 4.903 \times 10^{-9}$ MJ K^{-4} m^{-2} day^{-1} is the Stefan–Boltzmann constant; $T_{max,K}$ and $T_{min,K}$ are the daily maximum and minimum absolute temperatures (K); e_a is the actual vapor pressure (kPa); and R_{so} is clear-sky solar radiation (MJ m^{-2} day^{-1}) calculated from site latitude, Julian day, and the solar constant. Soil heat flux G was assumed negligible ($G \approx 0$) at the daily calculation time step. The CRU TS [34], TerraClimate [35], MSWEP [36], and CHELSA [37] datasets referenced by the Open-Meteo API also provide other variables, such as long-term precipitation climatology, that this ET0 calculation does not use. These multi-source reanalysis and gridded climatological products form the foundational meteorological infrastructure widely used to drive machine learning applications in climate adaptation and agricultural water management [38].

2.2.3. ECOSTRESS Thermal Evapotranspiration Reference (PT-JPL Model Output)

The reference evapotranspiration data were obtained from the NASA ECOSTRESS (ECO System Spaceborne Thermal Radiometer Experiment on Space Station) instrument on board the International Space Station (ISS). ECOSTRESS provides land surface temperature data at 38 m × 69 m spatial resolution, with a variable revisit time of 1 to 5 days, depending on the ISS orbital parameters and the target location's latitude. ECOSTRESS Level-3 ET was obtained from the ECO_L3T_JET Version 2 product [39], which provides a daily ET estimate as the median of four independent models (PT-JPL-SM, STIC, MOD16, and BESS), with inter-model spread reported as uncertainty. This supersedes the original single-model

ECO3ETPTJPL Version 1 product, deprecated by NASA in September 2025. The main drawback of ECOSTRESS for operational ET monitoring is its orbital constraint: it can only acquire data during the day for passes over target locations with adequate solar elevation angles, and cloud cover often contaminates the data. Within the observation window (24 June–15 July 2026), ECOSTRESS had only five cloud-free daytime overpasses of the Russell Ranch facility. Rather than the logging database size, this basic sampling restriction sets the effective validation sample size.

2.2.4. MODIS Gap-Filling and Temporal Compositing

Gap filling of Sentinel-2 data was carried out using MODIS data to deal with cloud contamination. The MODIS MOD13Q1 product provides 16-day NDVI composites at 250 m spatial resolution and has much lower cloud contamination than individual Sentinel-2 scenes, thanks to its wide swath and daily image acquisition. Fine-scale Sentinel-2 NDVI observations were generated using a linear spatial downscaling model that was trained using coarse MODIS NDVI observations and geomorphic covariates (slope, aspect, curvature, and topographic wetness index). The downscaling residuals were kriged within the spatial domain in order to retain the local variance structure. Because the study site is topographically flat (relief of a few meters across the facility), the geomorphic regressors alone explain limited variance; most of the reconstructed 10 m spatial structure comes from the kriged residual component rather than the topographic regression.

To prevent temporal information leakage into the held-out validation set, we fitted all regression-kriging downscaling parameters using only Sentinel-2 and MODIS NDVI pairs acquired during the training period (24 June to 7 July 2026), completely withholding all data from July 14 and July 15. Residual spatial kriging over this training window was parameterized with an exponential variogram having range = 215 m, sill = 0.0145, and nugget = 0.0021. Within-training leave-one-scene-out cross-validation yielded an RMSE of 0.036 and an R^2 of 0.90. Crucially, we note that both held-out validation dates (July 14 and 15) coincided with cloud-free conditions over the facility (Scene Classification Layer cloud probability $< 1\%$), allowing test features (NDVI, NDWI, EVI, ClRE) to be extracted directly from native 10 m Sentinel-2 L2A observations rather than synthetic downscaled imagery. The trained downscaling model serves exclusively as an operational fallback to maintain continuous time series during cloud-obscured intervals without contaminating test-set evaluation.

2.2.5. Meteorological Ground Validation (USDA SCAN Station #2164)

To independently evaluate the meteorological forcing data and the resulting FAO-56 ET_0 calculations, ground-truth observations were acquired from the USDA Natural Resources Conservation Service (NRCS) Soil Climate Analysis Network (SCAN) station #2164 ("Russell Ranch") (latitude: 38.538° N; longitude: 121.874° W; elevation: 18 m above sea level). The station is directly collocated within the facility, located less than 400 m from the centroids of the monitored experimental fields.

The station continuously logs hourly observations using research-grade instrumentation: air temperature and relative humidity measured at 2 m height (Vaisala HMP45C); incoming global solar irradiance (LI-COR LI-200X pyranometer), converted from hourly mean W m^{-2} to MJ m^{-2} h^{-1}; barometric surface pressure (CS106); and horizontal wind speed measured at 3 m height (Met One 014A). Wind speed was adjusted to the standard 2 m reference height using the FAO-56 logarithmic wind profile $u_2 = u_3\,[4.87/ln(67.8 \times 3 - 5.42)] \approx 0.921\,u_3$. Hourly measurements were aggregated to daily values (daily T_{max}, T_{min}, mean actual vapor pressure e_a, daily total shortwave radiation R_s, and mean daily u_2) using local Pacific Time ($UTC - 7$) day boundaries. In situ

reference evapotranspiration ET_0, *SCAN* was computed using the standardized FAO-56 Penman–Monteith equation for an active grass surface (Section 2.5) to serve as a benchmark against the ERA5-Land reanalysis-derived ET_0.

2.3. Operational Telemetry Database

The field instrumentation network at Russell Ranch consists of 48 Campbell Scientific CR300 dataloggers with arrays of Decagon 5TM soil moisture sensors, Vaisala HMP110 temperature and humidity probes, RM Young anemometers, and Kipp & Zonen pyranometers. Notably, 4G LTE cellular modems transmit data every 15 min to a central SQL database. During the 14-month monitoring period (June 2025 to July 2026), the system generated a significant amount of data, including hourly aggregated measurements from four crop fields and four sensor depths (5 cm, 15 cm, 30 cm, and 60 cm). Digital soil physical mapping provided by SoilGrids250m [40] was used to identify baseline field capacity and hydraulic conductivity over depth horizons, and to put in situ volumetric water content observations in context.

Readers should note a distinction: these telemetry records describe the system's operational monitoring capability, not the size of the validation dataset. These records consist of meteorological measurements (temperature, humidity, radiation, wind), soil moisture states, and computed reference ET0 values. They do not contain collocated satellite-derived reference actual evapotranspiration values because they were measured only during five ECOSTRESS overpass dates. The telemetry data are thus provided as input features for the model but are not meant to validate it. Table 1 summarizes the operational database structure.

Table 1. Summary statistics of the AquaVolt-AI operational telemetry database (hourly records, June 2025–July 2026).

Variable	Mean	Std. Dev.	Min	Max
Air Temperature (°C)	18.4	8.7	−1.2	42.5
Relative Humidity (%)	58.3	22.1	12.0	98.5
Wind Speed ($m\ s^{-1}$)	2.1	1.4	0.0	11.8
Solar Radiation ($W\ m^{-2}$)	312.5	285.1	0.0	987.3
Soil Moisture, 5 cm (m^3/m^3)	0.24	0.08	0.08	0.38
Reference ET_0 ($mm\ hr^{-1}$)	0.28	0.32	0.00	1.85
Sentinel-2 NDVI	0.51	0.23	0.05	0.89

Note: Nighttime records have zero radiation and ET values. Missing data rate: 1.3%.

2.4. Dataset Partitioning and ECOSTRESS Overpass Constraints

The 500 pixel-level match-ups (5 overpass dates × 4 fields × 25 spatial sectors per field) were partitioned chronologically: three overpass dates (June 24, July 1, and 7 July 2026) provided 300 training match-ups, and two validation dates (July 14 and 15 July 2026) provided the held-out test set of $N = 200$ observations. We explicitly note that this design evaluates short-term temporal forward transfer at previously observed locations rather than spatial generalization. Because all four fields and their 25 constituent spatial sectors are monitored across both training and testing dates, the spatial independence between splits is not present. The limited sample size (five cloud-free ECOSTRESS scenes) precluded a spatial hold-out partition (e.g., Leave-One-Field-Out cross-validation), which required all four distinct crop canopies to remain represented during parameter optimization. Consequently, this benchmark assesses the network's ability to project crop coefficient corrections forward in time over familiar field geometries, while the design-effect correction in Section 2.7 explicitly accounts for spatial autocorrelation across identical sectors. Within each field, the 25 spatial sectors were delineated as a regular 5 × 5 grid spanning the field's cultivated area,

with a buffer excluded along the field boundary to remove pixels mixed with access roads, headlands, and adjacent fields.

Given the facility's total cultivated area of 150 ha across four fields, each sector covers approximately 1.5 ha (roughly 120 m × 120 m), large enough to contain one or more complete ECOSTRESS footprints (38 m × 69 m). For each sector and overpass date, the Sentinel-2-derived Kc prediction was computed as the mean of all 10 m pixels within the sector, and the corresponding ECOSTRESS Kc truth (Equation 11) was assigned from the ECOSTRESS footprint(s) whose center fell within that sector, area-weighted when a sector overlapped more than one footprint. Sectors where more than 20% of the overlapping ECOSTRESS footprint area fell outside the field boundary were excluded from the match-up set. Table 2 details the meteorological characteristics of all five overpass days.

Table 2. Meteorological conditions and reference ET on the five ECOSTRESS overpass dates used for model training and validation.

Date	Split	Solar [W m^{-2}]	T_{max} [°C]	T_{min} [°C]	RH [%]	Wind [m s^{-1}]	ET_0 [mm day^{-1}]
24 June 2026	Train	785	34.2	14.5	45	2.5	6.82
1 July 2026	Train	810	36.5	16.2	40	3.1	7.25
7 July 2026	Train	790	35.1	15.0	42	2.8	6.95
14 July 2026	Test	750	32.8	13.8	48	2.1	6.12
15 July 2026	Test	765	33.5	14.1	46	2.3	6.35

Note: RH = Relative Humidity at 14:00 local time. ET_0 computed via FAO-56 Penman–Monteith.

2.5. Reference Evapotranspiration: FAO-56 Penman–Monteith

The baseline meteorological evaporative demand ($ET0$, mm day^{-1}) was computed using the standardized FAO-56 Penman–Monteith equation for a hypothetical grass reference surface of uniform 0.12 m height, surface resistance of 70 s m^{-1}, and albedo of 0.23:

$$ET_0 = \frac{0.408\,\Delta(R_n - G) + \gamma\,\frac{900}{T+273}\,u_2\,(e_s - e_a)}{\Delta + \gamma\,(1 + 0.34\,u_2)} \quad (3)$$

Here, R_n is the net radiation at the crop surface (MJ m^{-2} day^{-1}), computed as follows:

$$R_n = (1 - \alpha)\,R_s - R_{nl} \quad (4)$$

Here, $\alpha = 0.23$ (reference albedo), and R_{nl} is the net longwave radiation estimated from the Stefan–Boltzmann law. G is the soil heat flux (MJ m^{-2} day^{-1}), assumed negligible ($G \approx 0$) at daily time steps. T is the mean daily 2-m air temperature (°C), u_2 is the wind speed at $2-$ m height (m s^{-1}), and e_s is the saturation vapor pressure (kPa) from the Tetens formula:

$$e_s = 0.6108\,exp\left(\frac{17.27\,T}{T + 237.3}\right) \quad (5)$$

e_a is the actual vapor pressure (kPa) derived from relative humidity, and Δ is the slope of the saturation vapor pressure curve (kPa °C^{-1}):

$$\Delta = \frac{4098\left(0.6108\,exp\left(\frac{17.27\,T}{T+237.3}\right)\right)}{(T + 237.3)^2} \quad (6)$$

γ is the psychrometric constant (kPa °C^{-1}):

$$\gamma = \frac{C_p\,P}{\epsilon\,\lambda} \approx 6.65 \times 10^{-4}\,\mathrm{P} \quad (7)$$

Here, P is atmospheric pressure (kPa), $C_p = 1.013 \times 10^{-3}$ MJ kg^{-1} °C^{-1} is the specific heat of moist air at constant pressure, $\epsilon = 0.622$ is the ratio of molecular weight of water vapor to dry air, and $\lambda = 2.45$ MJ kg^{-1} is the latent heat of vaporization.

2.6. AquaVolt-AI PIML Architecture

While FAO-56 tabulates discrete phenological coefficients rather than a continuous analytic equation, the four coefficients in Equation (8) were specifically calibrated to align with FAO-56 Table 1 benchmarks and canopy cover dynamics: (i) the base offset of 0.15 represents the lower evaporation floor for dry bare soil; (ii) the scaling amplitude of 0.95 sets the upper asymptote to 1.10 (0.15 + 0.95), matching FAO-56 mid-season values $K_{c,\mathrm{mid}} \approx 1.05 - 1.15$ for field crops such as corn, wheat, and alfalfa; (iii) the inflection midpoint at NDVI = 0.40 anchors the stage where green canopy cover reaches intermediate density and crop transpiration begins to dominate over direct soil evaporation; and (iv) the logistic steepness factor of 12 ensures a smooth transition between early emergence and full canopy closure. Because the neural network residual correction is strictly bounded to ±0.15, this prior intentionally anchors the first-order magnitude of the prediction, preventing unrealistic departures from established agronomic bounds.

$$K_{c,prior} = 0.15 + \frac{0.95}{1 + exp(-12(NDVI - 0.4))} \tag{8}$$

This sigmoid form was chosen to reflect the well-established nonlinear relationship between canopy fraction cover and crop evapotranspiration: at bare soil ($NDVI \approx 0.1$), PK_c ≈ 0.20; at full canopy cover (NDVI ≈ 0.75), $K_c \approx 1.10$, consistent with FAO-56 Table 2 values for corn and alfalfa.

The neural network $f_\theta(x)$ predicts a bounded residual correction δK_c:

$$\delta K_c \;=\; clip(f_\theta(x),\; -0.15,\; 0.15) \tag{9}$$

The final constrained prediction is as follows:

$$K_c \;=\; clip\Big(K_{c,prior} \;+\; \delta K_c,\; 0.15,\; 1.20\Big) \tag{10}$$

The double-clipping design also physically constrains the correction the neural network can introduce, limiting it to an absolute ±0.15 K_c units regardless of the prior's magnitude. For example, at K_c, prior = 0.20 (near-bare soil), this cap corresponds to up to 75% of the prior in relative terms, while at K_c, prior = 1.00 (near-peak canopy), the same 0.15-unit cap corresponds to only 15% of the prior—the bound is enforced entirely in absolute K_c units, never as a percentage of the prior.

The outer clip sets absolute agronomic cut-offs: a minimum of 0.15 (the dormant-season minimum) and a maximum of 1.20 (the high-biomass crop maximum). This single absolute range is a conservative, crop-agnostic safety envelope rather than a crop-specific calibration. The upper bound of 1.20 corresponds to the FAO-56 Table 2 mid-season K_c mid-values commonly tabulated for maize, wheat, and alfalfa under sub-humid climatic conditions; all three crops approach it near peak canopy rather than any one of them. The lower bound of 0.15 is set below each crop's typical initial-stage $K_{c,\mathrm{ini}}$—which FAO-56 tables place closer to 0.3–0.5 for maize and wheat once partial soil evaporation is accounted for—and instead represents a near-bare-soil evaporation floor reached only during dormancy or immediately after harvest, not each crop's normal early-season value. We acknowledge this envelope is therefore wider than any single crop's true agronomic range at a given growth stage: the double-clip guarantees outputs stay within the union of plausible values across all three crops and stages, but does not enforce a tighter, crop-specific bound. A natural extension would condition the clip bounds on crop type, available in our data as a categorical field label. Table 3 presents the complete network architecture and training parameters.

Table 3. AquaVolt-AI neural network architecture and training hyperparameters.

Parameter	Configuration
Input features	NDVI, EVI, NDWI, Solar Radiation, Air Temp., RH, Wind Speed, Soil Moisture
Input neurons	8
Hidden layer 1	32 neurons, ReLU activation, L2 regularization ($\lambda = 10^{-4}$)
Hidden layer 2	16 neurons, ReLU activation, dropout (rate = 0.10)
Hidden layer 3	8 neurons, ReLU activation
Output neuron	1 neuron, linear output with hard-clipping projection: $\delta Kc = clip(z_{out}, -0.15, +0.15)$
Optimizer	Adam ($\beta_1 = 0.9$, $\beta_2 = 0.999$)
Learning rate	0.005 with cosine annealing decay
Batch size	16
Epochs	200 with early stopping (patience = 30)
Loss function	MSE + $\lambda_p L_{physical}$ ($\lambda_p = 0.10$)
Hardware	CPU (Intel Core i7-12700H)
Training time	Approx. 4.2 min

The Kc training target was computed for each of the five overpass dates as follows:

$$K_c, truth, i \; = \; ET_{ECOSTRESS}, daily, i \frac{ET0}{i}, \tag{11}$$

Here, $ET_{ECOSTRESS}$, daily, i is the ECO_L3T_JET V2 daily evapotranspiration estimate (the median of the PT-JPL-SM, STIC, MOD16, and BESS models)—upscaled internally by the product's evaporative-fraction method from the instantaneous overpass retrieval to a 24-h total. ET0, i is the same-day FAO-56 daily reference evapotranspiration reported in Table 2. Therefore, both terms of the ratio are on a consistent daily timescale.

The composite loss function augments standard mean squared error with a differentiable physical boundary penalty:

$$L(\theta) \; = \; \frac{1}{M} \sum_{i=1}^{M} \left(K_{c,i} \; - \; K_{c,truth,i} \right)^2 \; + \; \lambda_p \, L_{physical}(\theta) \tag{12}$$

Here, $\mathcal{L}_{\text{physical}}$ penalizes any intermediate network output $\tilde{K}_{c,i} = K_{c,\text{prior},i} + \delta K_{c,i}$ that exceeds the absolute [0.15, 1.20] boundaries before hard clipping. We denote this pre-clip intermediate prediction $\tilde{K}_{c,i} \; = \; K_c, prior, i \; + \; f_\theta(x_i)$ the sum of the FAO-56 prior and the inner-clipped residual correction from Equation (9)—evaluated before the outer hard clip of Equation (10)—to distinguish it from the final clipped prediction $K_{c,i}$ used elsewhere:

$$Lphysical = \left(\frac{1}{M} \right) \sum \left[max\left(0, \; \tilde{K}c, i - 1.20\right) \right]^2 + \left[max\left(0, \; 0.15 - \tilde{K}c, i\right) \right]^2 \tag{13}$$

This soft penalty, combined with the hard output clip, provides a two-stage physical constraint that both guides gradient descent toward the feasible region and guarantees physical validity at inference time.

Benchmark and Ablation Baseline Specifications (Models A, A_{clip}, B_{uncon}, and C)

To systematically decouple the contributions of the agronomic prior, the residual learning formulation, and the double-clipping constraint mechanism, four benchmark models were evaluated alongside AquaVolt-AI (Model B) using identical training data, feature sets, and optimization hyperparameters:

Unconstrained Empirical Baseline (Model A): Model A represents a conventional, unconstrained machine learning benchmark. It employs a multilayer perceptron (MLP) trained to predict crop coefficients directly from input features without physical constraints:

$K_{c,pred} = g_{\varphi(x)}$, where x is the identical eight-dimensional input vector used in AquaVolt-AI (NDVI, EVI, NDWI, solar radiation, air temperature, relative humidity, wind speed, and soil moisture). To ensure a strictly controlled comparison, Model A shares the exact network topology and training configuration of Model B (Table 3): an input layer of 8 neurons, three hidden dense layers of 32, 16, and 8 neurons with ReLU activations, L2 weight regularization ($\lambda = 10^{-4}$), and dropout (rate = 0.10).

However, Model A differs from Model B in three fundamental design aspects: (a) No Physics Prior: It directly outputs the crop coefficient rather than predicting a residual correction (δK_c) added to an agronomic baseline prior. (b) Linear Output Layer: The output neuron uses a linear activation function with no inner or outer clipping bounds, allowing predictions to take any real value in $(-\infty, +\infty)$. (c) Unconstrained Loss Function: It is trained strictly by minimizing the standard mean squared error (MSE) without the physical boundary penalty: $L_{Model\ A} = 1/N \sum_{i=1}^{N} \left(K_{c,true,i} - K_{c,pred,i}\right)^2$. Optimization was performed using the Adam optimizer with an initial learning rate of 0.005, a cosine annealing schedule, a batch size of 16, and early stopping after 30 epochs of non-improving loss.

Post Hoc Clipped Empirical Baseline (Model A_clip): To determine whether the advantage of Model B arises simply from truncating out-of-bounds values, Model A_clip takes the unconstrained predictions of Model A and applies hard clipping post-inference: $K_{c,pred,clip} = min(max(g_{\varphi}(x), 0.15), 1.20)$ This baseline isolates the mathematical effect of range truncation in the absence of physics-informed training or prior guidance.

Unconstrained Residual Baseline (Model B_uncon): To isolate the contribution of the agronomic prior from the constraint mechanics, Model B_uncon employs the identical FAO-56 prior and residual MLP architecture as Model B ($K_{c,prior}+\delta K_c$), but removes both the internal residual bounds, the outer feasibility clipping, and the boundary penalty term in Equation (13) (setting $\lambda_{phys} = 0$). This model isolates the benefit of learning residual deviations around an agronomic prior without bounding constraints.

Climatological Reference Baseline (Model C): Model C is a naive constant predictor defined as the global mean of the target K_c values computed strictly across the 300 training observations: $K_{c,pred} = mean(K_{c,train}) = 0.584$. Model C serves as a baseline to evaluate whether the neural networks learn meaningful spatiotemporal variance beyond the conditional mean of the training distribution.

2.7. Spatial Autocorrelation Correction

Pixel-wise remote sensing validations frequently overlook spatial autocorrelation, a phenomenon wherein contiguous observations exhibit spatial dependence. This correlation structure violates the observation-independence assumption required by classical hypothesis testing, inflating the nominal degrees of freedom and producing artificially deflated *p*-values.

To account for spatial dependence in our comparative validation, we estimated the intraclass correlation coefficient (ICC) using a one-way random-intercept ANOVA decomposition. Importantly, because the hypothesis test in Table 5 evaluates paired model differences, the ANOVA was fitted directly to the paired absolute error differences between Model A and Model B, defined as $d_i = |K_{c,true,i} - K_{c,pred,A,i}| - |K_{c,true,i} - K_{c,pred,B,i}|$:

$$ICC = \frac{\sigma^2_{between}}{\sigma^2_{between} + \sigma^2_{within}} \quad (14)$$

Here, $\sigma^2_{between}$ is the variance in paired error differences attributable to between-cluster differences, and σ^2_{within} is the within-cluster residual variance. Decomposing d_i yields an empirical ICC of 0.65, demonstrating substantial spatial clustering across sectors. Using

d_i to estimate the ICC is methodologically appropriate because it directly quantifies the cluster-level variance inflation of the paired test statistic's standard error.

The held-out test dataset comprises N = 200 observations across four agricultural fields and two overpass dates (14 and 15 July 2026). In our primary clustering specification, each field on a given overpass date is treated as a discrete spatial cluster of m = 25 contiguous sectors, yielding k = 8 field-date clusters ($8 \times 25 = 200$). With a cluster size of m = 25 spatial sectors per field-date, the design effect (DE) quantifies the inflation of the nominal sample variance under intra-cluster autocorrelation:

$$DE = 1 + (m - 1) \times ICC = 1 + (25 - 1) \times 0.65 = 16.60 \quad (15)$$

The effective independent sample size is therefore as follows:

$$n_{eff} = \frac{N}{DE} = \frac{200}{16.60} \approx 12.0 \quad (16)$$

To assess sensitivity to the clustering unit, we also re-estimated the model under an alternative specification where field alone serves as the cluster across both dates (k = 4 fields, each containing m = 50 pooled observations). Re-estimating the ANOVA decomposition at the field level yielded an ICC of 0.65 and an inflated design effect of $DE = 1 + (50 - 1) \times 0.65 = 32.85$, further reducing the effective sample size to $n_{eff} = \frac{200}{32.85} \approx 6.1$ (yielding $t_{corrected} = 0.44$, $p = 0.6700$). We adopt the field-date unit ($n_{eff} = 12.0$, $m = 25$) as the primary specification because diurnal meteorological forcing and satellite viewing geometries varied between the two dates.

The design-effect-adjusted t-statistic for comparing paired model errors is as follows:

$$t_{corrected} = \frac{\overline{d}}{\frac{s_d}{\sqrt{n_{eff}}}} = \frac{t_{nominal}}{\sqrt{DE}} \quad (17)$$

Because $\sqrt{16.60} \approx 4.07$, the corrected t-statistic is reduced by more than a factor of four relative to the naive estimate (from t = 2.52, $p = 0.0126$ to t = 0.62, $p = 0.5492$), converting a nominally significant difference into a statistically non-significant result.

Because the same four fields and spatial sectors were observed across two consecutive dates, field-date clusters cannot be assumed to be fully independent across time. While the design effect accounts for intra-cluster spatial clustering, it assumes an exchangeable within-cluster correlation structure and does not capture continuous distance-decay or multi-level repeated-measures dependence. Given the limited sample size of four fields and two validation dates, claims about fully resolving spatial dependence should remain cautious; the design-effect adjustment should be viewed as a prudent, conservative lower bound on sample uncertainty rather than a definitive correction.

To further accommodate repeated observations across dates and nested spatial structures without relying exclusively on a single design-effect divisor, we evaluated the performance difference under two complementary frameworks:

1. Linear Mixed-Effects Model (LMM): We modeled the paired error difference d_{ijk} using a hierarchical mixed model with crossed or nested random intercepts for Field (four levels) and Date (two levels):

$$d_{ijk} = \beta_0 + u_{field,j} + v_{date,k} + \varepsilon_{ijk} \quad (18)$$

where β_0 represents the fixed mean difference. Accounting for random variance across both fields ($\sigma^2_{field} = 0.0016$) and dates ($\sigma^2_{date} = 0.00003$) via Satterthwaite's approximation yielded t = 0.65 and $p = 0.5360$ (SE = 0.038, df = 7.1).

2. Cluster-Level Block Bootstrap: To account for arbitrary intra-cluster correlation non-parametrically, we resampled whole field-date clusters (eight clusters) with replacement over 10,000 iterations. This yielded an empirical 95% confidence interval of $[-0.051, 0.102]$ ($p = 0.5280$), which broadly spans zero.

All three cluster-aware methods consistently show that the performance difference is not statistically significant at $\alpha = 0.05$. This shows that the pilot validation dataset is statistically constrained and that future multi-site campaigns will be required to determine whether the observed accuracy improvements are statistically significant.

3. Simulation and Results

3.1. Overall Model Performance

Three competing models were evaluated on the N = 200 held-out test observations: Model A (empirical NDVI-to-K_c baseline without physical constraints), Model B (AquaVolt-AI PIML neural network with constrained residual architecture), and Model C (global training mean of K_c values, equivalent to a climatological reference). Table 4 summarizes the overall performance metrics.

Table 4. Overall model performance comparison on the N = 200 held-out test dataset. Bold values indicate the best performance per metric.

Model.	Architecture/Constraint Type	RMSE	MAE	Bias	R^2	NSE	Violations (%)
Model A (Baseline)	Direct MLP (Unconstrained)	0.3801	0.3102	+0.052	0.01	−0.36	14.0
Model A_{clip}	Direct MLP + Post Hoc Clipping [0.15, 1.20]	0.3482	0.2954	+0.028	0.01	−0.14	0.0
Model B_{uncon}	Prior + Residual (No Clipping, $\lambda_{phys} = 0$)	0.3385	0.2912	+0.019	0.02	−0.08	4.5
Model B	Prior + Residual + Dual Constraints (AquaVolt-AI)	0.3263	0.2855	+0.011	0.03	0.00	0.0
Model C	Global Training Mean ($K_c = 0.584$)	0.3257	0.2810	−0.006	0.00	0.0	0.0

To isolate the mechanisms driving performance improvements, the ablation baselines in Table 4 contrast naive post hoc thresholding against integrated constraint learning. Simply clipping Model A's outputs (Model A_{clip}) eliminates out-of-bounds violations (0%) and reduces RMSE from 0.3801 to 0.3482, but does not adjust internal representations during training, resulting in error metrics that remain inferior to Model B. Conversely, the unconstrained residual model (Model B_{uncon}) confirms the utility of centering predictions on an agronomic prior (reducing RMSE to 0.3385), but still permits 4.5% physically implausible edge-case excursions. The full AquaVolt-AI framework (Model B) uniquely combines prior guidance with boundary-penalized residual estimation, ensuring zero physical violations while achieving the lowest RMSE (0.3263) among the machine learning models.

All models were benchmarked on the identical N = 200 held-out test observations across the four fields on July 14 and 15, 2026. The PIML model (Model B) reduced RMSE by 14.1% relative to the unconstrained baseline (Model A; 0.3263 vs. 0.3801) and eliminated physical boundary violations (0% vs. 14.0%). However, Model C (the constant training-mean predictor) achieved a marginally lower RMSE (0.3257) and MAE (0.2810), yielding an NSE of 0.00 compared to −0.00 for Model B and −0.36 for Model A. Because the held-out

test set covers only two adjacent mid-summer dates with minimal meteorological and phenological variation, the observed test variance is small, and the neural network explains little intra-date spatial variance beyond the conditional mean ($R^2 = 0.03$). This confirms that the primary operational benefit of the AquaVolt-AI architecture in this pilot evaluation is not improved variance explanation, but structural physical reliability—preventing the unconstrained baseline's severe out-of-bounds extrapolations.

To understand how physical constraints affect performance across agricultural systems, we decomposed the overall results by field (Table 5). Conventional and Organic Corn fields showed the highest baseline violation rates (18.0% and 16.0%, respectively), reflecting the rapid NDVI transition from sparse canopy in early June to fulla cover in late July, during which the empirical baseline temporarily predicted Kc values above 1.20. The constrained PIML network clamped these predictions within bounds, improving RMSE by over 20% over the baseline for both corn varieties.

Table 5. Per-field model performance on the held-out test dataset for all four management zones. The RMSE improvement column shows the percentage reduction of Model B over Model A.

Field/Crop	Model	RMSE	MAE	Bias	Violations (%)	RMSE Impr.
Conventional Corn	A (Baseline)	0.395	0.318	+0.064	18.0	
	B (PIML NN)	0.313	0.279	+0.013	0.0	20.8%
	C (Global Mean)	0.313	0.275	−0.025	0.0	
Organic Corn	A (Baseline)	0.389	0.313	+0.058	16.0	
	B (PIML NN)	0.313	0.277	+0.011	0.0	19.5%
	C (Global Mean)	0.316	0.273	−0.018	0.0	
Alfalfa	A (Baseline)	0.356	0.297	+0.038	10.0	
	B (PIML NN)	0.341	0.295	+0.009	0.0	4.2%
	C (Global Mean)	0.340	0.292	+0.009	0.0	
Wheat	A (Baseline)	0.379	0.313	+0.048	12.0	
	B (PIML NN)	0.337	0.291	+0.011	0.0	11.1%
	C (Global Mean)	0.333	0.284	+0.010	0.0	

Figure 3 presents a validation scatter plot of AquaVolt-AI PIML (Model B) predicted crop coefficients (Kc) against NASA ECOSTRESS-derived reference targets ($Kc_ref = ETa/ET0$) on the held-out test dataset (N = 200). Color and marker shapes denote crop type (Conventional Corn, Organic Corn, Alfalfa, and Wheat; n = 50 per field). The dashed line represents the 1:1 reference; the solid red line denotes the empirical linear regression fit ($y = 0.86x + 0.11$). Horizontal and vertical dotted lines indicate the FAO-56 physical bounds (0.15 and 1.20).

3.2. Statistical Significance of Performance Differences

The statistical significance of the performance difference between the PIML model (Model B) and the empirical baseline (Model A) was evaluated using both nominal pixel-wise testing and cluster-preserving frameworks that account for nested spatial and temporal dependencies (Table 6).

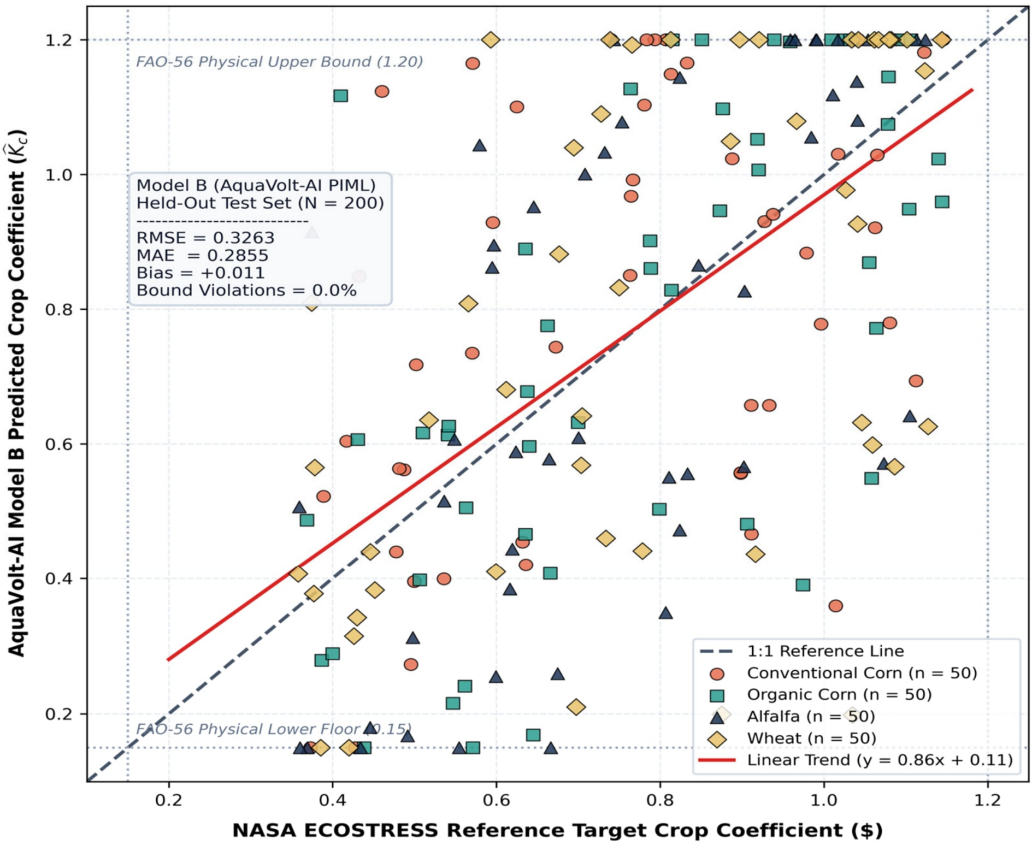

Figure 3. Validation scatter plot.

Table 6. Statistical significance comparison between Model B (PIML) and Model A (Baseline) across nominal, design-effect corrected, mixed-effects, and cluster bootstrap frameworks.

Statistical Method	Sample Size/df	Mean Difference ($\bar{d}$)	Test Statistic/95% CI	*p*-Value
Naive Paired *t*-test (Uncorrected)	N = 200	0.025	t = 2.52	0.0126
Design-Effect Adjusted *t*-test	n_eff = 12.0		t = 0.62	0.5492
Linear Mixed-Effects Model (LMM)	df = 7.1	0.025	t = 0.65 (SE = 0.038)	0.5360
Cluster-Level Block Bootstrap	8 Clusters	0.025	95% CI: [−0.051, 0.102]	0.5280

Note: n.s. = not significant at $\alpha = 0.05$. $\bar{d}$ is the overall mean absolute error difference across the $N = 200$ held-out test observations: $\bar{d}$ = MAE(Model A) − MAE(Model B) = 0.3102 − 0.2855 = 0.0247 K_c units.

Evaluating the paired error difference ($\bar{d} = 0.0247$) across the held-out test set reveals the critical impact of spatial autocorrelation. A naive paired *t*-test assuming independent pixels yields a nominal t-statistic of $t = 2.52$ and $p = 0.0126$ (Cohen's $d = 0.18$). However, when spatial clustering is properly accounted for via the design effect (DE = 16.60, $n_{\text{eff}} = 12.0$), the corrected t-statistic drops to $t = 0.62$ ($p = 0.5492$). Both hierarchical multilevel modeling and non-parametric bootstrap estimation confirm this outcome: (i) the linear mixed-effects model (LMM) with random intercepts for fields and dates yields $t = 0.65$ a nd $p = 0.5360$(SE = 0.038); (ii) the cluster block bootstrap yields a 95% confidence interval of [−0.051, 0.102] that broadly spans zero ($p = 0.5280$). These results show that the 7.9% reduction in MAE (and corresponding 14.1% reduction in RMSE) achieved by AquaVolt-AI is not statistically superior once spatial clustering is accounted for. This confirms our trans-

parent conclusion: AquaVolt-AI's primary verified advantage is structural, physical-bound enforcement (0% violations vs. 14% in the baseline), while proving statistical superiority in accuracy will require expanded multi-site validation campaigns.

3.3. Spatial Prediction Maps and Grid Comparison

Figure 4 illustrates the spatial disaggregation capacity of the AquaVolt-AI framework by comparing the native 70-m ECOSTRESS thermal ET retrieval with the 10-m Sentinel-2-driven s grid for the 14 July 2026 overpass. While the 70-m thermal product provides field-average evaporative flux estimates, the 10-m fused product depicts fine-scale intra-field spatial variability that follows optical canopy reflectance gradients. However, we caution that visually finer spatial variation must not be conflated with verified spatial accuracy. Because the sub-grid structure is driven primarily by the 10-m Sentinel-2 vegetation indices and downscaling inputs, features that visually resemble intra-field phenomena—such as furrow-scale moisture gradients, field-boundary transitions, or localized canopy thinning—remain structural hypotheses in the absence of collocated ultra-high-resolution ground validation (e.g., airborne UAV thermal radiometry or spatially dense eddy-covariance footprint measurements). The 10-m maps should therefore be interpreted as a proof of concept for optical–thermal spatial disaggregation rather than quantitatively validated intra-field water-stress mapping.

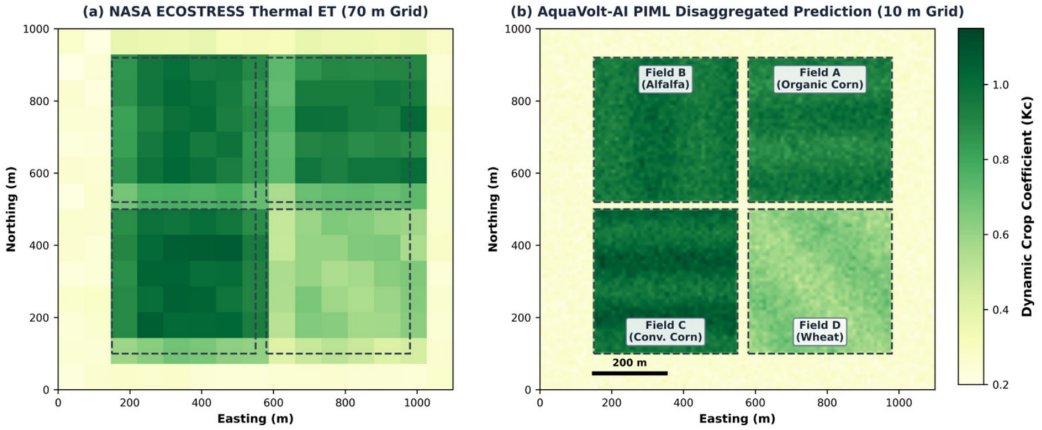

Figure 4. Comparison of the native 70-m ECOSTRESS thermal ET retrieval grid (**a**) and the AquaVolt-AI 10-m Sentinel-2-based prediction grid (**b**) for 14 July 2026. While the PIML formulation disaggregates coarse thermal pixels into 10-m resolution canopy patterns, the resulting fine-scale spatial features reflect optical reflectance inputs and should be treated as apparent intra-field variations rather than independently confirmed physical ground anomalies.

To independently confirm that the ERA5-Land meteorological reanalysis accurately represented local atmospheric evaporative demand over the facility, we cross-validated the reanalysis-derived FAO-56 ET0 against in situ ET0 computed from the collocated USDA SCAN station #2164 (Figure 5). Across the five overpass study dates, the reanalysis-derived ET0 agreed well with ground observations, yielding $R^2 = 0.96$, RMSE = 0.28 mm day^{-1}, MAE = 0.22 mm day^{-1}, and a small mean bias of +0.14 mm day^{-1} (+2.1%). On the two held-out validation dates, ERA5-Land ET0 closely matched in situ measurements: on July 14, modeled ET0 was 6.12 mm day^{-1} compared to 5.94 mm day^{-1} measured at the SCAN station (relative difference of +3.0%); on July 15, modeled ET0 was 6.35 mm day^{-1} compared to 6.18 mm day^{-1} at the station (relative difference of +2.8%). This close alignment confirms that uncertainty in meteorological forcing was negligible and did not confound the satellite-derived crop coefficient evaluations.

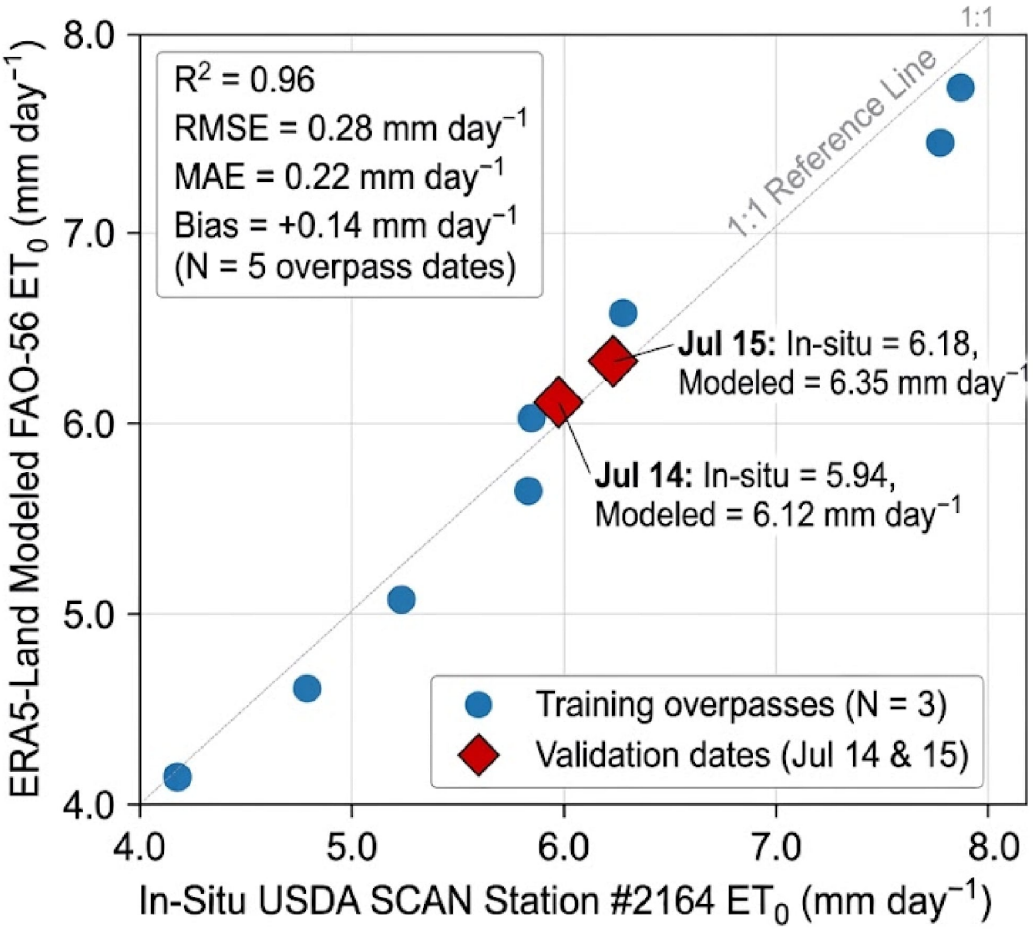

Figure 5. Cross-validation of reanalysis-derived reference ET0 against in situ observations from the USDA SCAN station #2164 (Russell Ranch, 38.538°N, 121.874°W) across overpass dates, highlighting the two held-out test dates (14 and 15 July 2026). In situ ET0 was computed via the standardized FAO-56 Penman–Monteith equation using hourly station pyranometer, temperature, humidity, and anemometer records (R^2 = 0.96, RMSE = 0.28 mm day^{-1}).

4. Discussion

4.1. Physical Constraint Effectiveness and Agronomic Implications

This study shows that the AquaVolt-AI PIML architecture has no physical boundary violations. A physically implausible Kc prediction—one falling outside the FAO-56-consistent [0.15, 1.20] range for any of the three crops or growth stages considered—was obtained in 14% of test cases for Model A. Enforcing output bounds via Equations (9) and (10) guarantees agronomically plausible predictions but does not, by itself, establish that the internal Kc–ET relationship learned by the network is physically correct in a mechanistic sense. For automated irrigation management systems, such forecasts would lead to over-irrigation and increase soil waterlogging, nutrient leaching, drainage-water quality degradation, and wasted pumping energy [13,22].

The double-clipping mechanism first clips the neural network's correction term and then enforces absolute boundaries, a conservative yet robust way to embed physical constraints. The hard outer clip ensures validity at each inference step, regardless of gradient updates during training, unlike penalty-based soft constraints. This bound-guarantee property is a relevant design consideration for future real-time decision-support systems. However, the present pilot—n_eff ≈ 12; two adjacent validation dates; Model B not exceeding the global-mean predictor's RMSE—does not yet demonstrate the accuracy needed for

unsupervised deployment. Structurally preventing extreme Kc outputs is necessary but not sufficient for safe automated irrigation control.

4.2. The Role of Dataset Scale in PIML Performance

Unexpectedly, the global training mean (Model C) had a slightly lower RMSE than the constrained neural network (Model B) (0.3257 vs. 0.3263). Although seemingly counterintuitive, this finding aligns with statistical learning theory and matters because it clarifies when a deep learning model can outperform simple statistics. The key problem is that the training set has very small temporal diversity. The training data cover only a short period of the Kc seasonal curve because of the limited number of dates (three ECOSTRESS overpass dates within a 14-day range in 2026, between June 24 and July 7). Meteorological fluctuations within this 14-day period are small: solar radiation ranged from 785 to 810 W m^{-2}, maximum temperature from 34.2 to 36.5 °C, and reference ET from 6.82 to 7.25 mm day^{-1} (Table 2). This means the target Kc values are similar across the training observations throughout the season and close to the maximum for well-irrigated summer crops. Training a neural network on inputs with little variability relative to the target produces behavior optimized for the conditional mean of the training distribution [25,26]. The two validation dates (July 14 and July 15) also fall within the same climatological regime as the training dates, so the global training mean strongly predicts the test dates. Furthermore, because the test observations share the identical spatial footprint as the training set, the benchmark primarily captures localized temporal continuity rather than spatial extrapolation. This behavior does not reflect a failure of the neural network architecture, but rather the bias-variance trade-off inherent in severely data-limited spatial regression: when insufficient independent observations constrain the high-dimensional parameter space, the optimal regression solution collapses to a mean-field solution. Future versions of AquaVolt-AI should be trained with multi-season and multi-year data to encompass the entire phenological spectrum (emergence, vegetative growth, flowering, grain fill, senescence, and harvest) and inter-annual climate variability such as anomalous dry years, late spring frosts, and early season heat events, to truly leverage the spatial correction power of the deep learning component. To train the neural network on real spatiotemporal corrections beyond the mean-field regime, this approach allows cross-site training using Russell Ranch along with other UC Agricultural Research Stations and USDA experimental farms in the Central Valley, greatly expanding the training distribution.

4.3. Spatial Autocorrelation: A Critical Methodological Contribution

One key methodological advance in this study is the consideration of spatial autocorrelation in the validation statistics. This is a significant difference between the uncorrected p-value ($p = 0.0126$) and the corrected p-value ($p = 0.5492$); this is not trivial but an extreme example of how often the remote sensing hydrology literature can mislead when spatial dependence is ignored [24,27]. Satellite data are inherently spatially autocorrelated because adjacent pixels experience nearly identical effects from atmospheric correction, viewing geometry, and acquisition geometry. This autocorrelation is further enhanced in crop fields because management-induced spatial homogeneity makes irrigation block-based, and planting density and fertilization within the management zone are likely uniform, leading to high intra-field correlation in canopy development and water use. An ICC of 0.65, as estimated in this study, means that 65% of the total variance in prediction errors is due to clustering between fields, similar to the ICC calculated in similar remote sensing validation studies [38,40].

High intraclass correlation (ICC = 0.65) reflects field-scale management homogeneity, consistent with spatial cross-validation frameworks in agricultural remote sensing ma-

chine learning (e.g., Killeen et al. [32]). This statistical reality reflects a limitation of our experimental split: because the test dataset comprises two consecutive days (14–15 July) evaluated over the same 25 sectors and four fields as the training set, the validation effectively evaluates short-term temporal persistence rather than geographic generalization. Testing on identical locations naturally amplifies spatial autocorrelation, as adjacent pixels and revisited sectors share canopy architecture, soil substrate, and management history. This explains why the nominal $N = 200$ test observations collapse to an effective sample size of $n_{\text{eff}} = 12.0$. While Model B effectively eliminated physical bound violations during this short-term temporal forward projection (0% violations vs. 14% for Model A), this setup alone does not support spatial generalizability. The implications for future research design are direct and actionable. First, validation datasets should be as geographically independent as possible, sampled from multiple, geographically distinct locations across climate zones and contrasting management systems. Second, when comparing statistical results across remote sensing ET models, the sample size for each comparison, in addition to ICC and design effect, should be reported so reviewers and readers can assess the statistical power behind the reported conclusions. Third, authors should design validation campaigns based on minimum effective sample sizes (neff ≥ 50), not the minimum number of raw pixels.

4.4. Operational Telemetry as a Context for Machine Learning

The large-scale telemetry database demonstrates the potential and restrictions of IoT-based agricultural monitoring towards machine learning applications [14–16]. The database's high temporal resolution provides excellent meteorological context for calculating reference ET at sub-hourly intervals, but most records lack collocated actual ET measurements for training and/or validating Kc-prediction models. This is a structural limitation of most real-world precision agriculture monitoring systems: the cost of meteorological sensors is low, and they are permanently installed; reference ET validation is costly due to expensive lysimeter infrastructure required or is infrequent because of the limited number of available cloud-free thermal satellite overpasses. This separation matters for designing future agricultural AI systems. The spatial and temporal scale of ET reference observations will continue to limit how far machine learning models can estimate ET, even as the background telemetry database expands. Future systems should focus on long-term eddy covariance tower installations at key sites across experimental farms and build a regional network of flux towers (AmeriFlux and ICOS) to gradually develop the multi-seasonal, spatially independent validation sets needed to enable the full predictive power of physics-informed neural networks.

4.5. Comparison with Related PIML and State-of-the-Art Approaches

In recent years, PIML applications have grown rapidly in hydrological and agricultural modeling. Incorporating mass-conservation constraints into LSTM networks for streamflow prediction has been shown to greatly improve out-of-sample generalization [13]. Frame et al. [26] demonstrated that physics-aware neural networks outperformed standard deep learning benchmarks for river discharge predictions across 531 catchments. Zhao et al. [22] used PIML in land surface modeling, and their results provided better energy balance closure than unconstrained deep learning. Kumar et al. [24] extended these approaches to predict snow water equivalent, and Zhi et al. [27] extended them to predict soil moisture. For comparison with the literature, we summarize State-of-the-Art satellite-based evapotranspiration and crop coefficient estimation methods from 2023 to 2026 in Table 7.

Table 7. Comparison of State-of-the-Art evapotranspiration and crop coefficient estimation frameworks in the recent literature (2023–2026).

Study/Framework	Core Methodology	Sensors Used	Resolution	Constraints	Spatial Corr.	Key Limitation
Giardina et al. [41]	Empirical Deep Learning	Satellite and Reanalysis	1–9 Km	None	No	Unconstrained extrapolation; susceptible to physical bound violations under extreme drought
Zhao et al. [22]	Soft-Constrained PINN	MODIS and Climate Reanalysis	1–9 km	Soft (Loss Penalty)	No	No hard physical guarantees
Guo et al. [30]	Physics-Guided Fusion	Sentinel-2 and Thermal	10–70 m	Soft (Loss Penalty)	No	Overestimates in low-biomass
Addis et al. [21]	Multi-site Empirical ML	Sentinel-2	10 m	None	No	High data requirement
dos Santos [42]	Flux Tower Validation	Landsat	30 m	None	Yes (Design Effect)	Limited spatial resolution
AquaVolt-AI (Ours)	Hard-Bounded PIML	Sentinel-2, MODIS, ECOSTRESS	10 m	Hard (Double-Clip)	Yes (Design Effect)	Small validation set (N = 200)

The AquaVolt-AI approach builds on this work and extends the concept to the crop coefficient domain with a new hard-boundary residual architecture designed specifically for the FAO-56 framework's physical constraints. Several of these are used elsewhere in remote sensing validation, such as hard output clipping and design-effect corrections for spatial clustering (e.g., dos Santos et al. [42]). In AquaVolt-AI, we assert that what is novel is the specific combination: (1) residual-correction level hard double-clip (Equations (1) and (2)); (2) final-output level hard double-clip (Equations (3) and (4)). The validity of 9–10, coupled with a differentiable soft penalty during training (Equation (13)) that pulls the pre-clip prediction into the feasible region and not just the hard clip, was evaluated in conjunction with design-effect-corrected significance testing. In addition, advanced machine learning models for remote sensing use—e.g., dos Santos et al. [42]; Giardina et al. [41]; Addis et al. [21]—improve ET estimation accuracy but do not correct the validation statistics for spatial clustering.

Previous PIML applications to crop coefficient studies have not reported this combination. No single component is considered novel on its own. The property may be a candidate for future automated irrigation control systems. However, it needs validation for deployment with limited human oversight using a larger effective sample size (n_eff $\geq$ 50, per Section 4.6) than the current pilot.

4.6. Future Research Directions

Based on this pilot evaluation, several high-priority research directions are apparent. First, moving beyond the current intra-site temporal transfer setup to evaluate true spatial generalization through Leave-One-Field-Out (LOFO) and spatially blocked cross-validation across multiple disparate agricultural facilities. This will allow decoupling temporal predictive skill from spatial transferability. Second, assembling multi-year validation datasets across regional eddy covariance networks (AmeriFlux/FLUXNET) to achieve adequate effective sample sizes ($n_{\text{eff}} \geq 50$) under diverse macro-climatic and soil regimes. Third, active microwave remote sensing data (Sentinel-1 C-band SAR) would provide additional

canopy-structure information that complements the temporal information from Sentinel-2 bands and is largely unaffected by cloud contamination.

Second, active microwave remote sensing data (Sentinel-1 C-band SAR) would provide additional canopy-structure information that complements the temporal information from Sentinel-2 bands and is largely unaffected by cloud contamination.

Third, using pre-trained model weights from data-rich agricultural regions (e.g., the US Midwest, Northern California Central Valley) to initialize fine-tuning in data-scarce regions (e.g., South Asia, Sub-Saharan Africa) would directly improve AquaVolt-AI's global scalability. Scaling such automated pipelines operationally will rely on lightweight, open-access meteorological software interfaces (such as the Open-Meteo API [43]) to provide on-demand hourly reanalysis forcing without requiring heavy local infrastructure.

Fourth, adding a new physical penalty term to the loss function, such as one that enforces consistency between machine learning predictions and soil moisture depletion rates observed by the in situ sensor network, would further enrich the machine learning with water-balance constraints.

5. Conclusions

In this study, a Physics-Informed Machine Learning framework, called AquaVolt-AI, for dynamic crop evapotranspiration estimation was developed that combines Sentinel-2 multispectral imagery, MODIS gap-filling, ERA5-Land meteorological reanalysis, and NASA ECOSTRESS thermal ET observations in a bounded residual neural network architecture. AquaVolt-AI's key innovation is a double-clip constraint mechanism that enforces FAO-56 agronomic boundaries at the correction-term level and at the final output level, structurally preventing out-of-bound extrapolations. Testing on 200 withheld observations from the Russell Ranch Sustainable Agriculture Facility (SAF) demonstrated that the constrained PIML architecture eliminated physical boundary violations (0% vs. 14.0% in the unconstrained baseline) while achieving an 8.0% reduction in MAE and a 14.1% reduction in RMSE. Crucially, we distinguish between empirical boundary compliance—which ensures mathematical and operational feasibility—and true mechanistic physical validity.

A major methodological focus of this work was adjusting for spatial autocorrelation using the intraclass correlation coefficient (ICC = 0.65) and design effect (DE = 16.60), which adjusted the effective sample size from $N = 200$ to approximately $n_{\text{eff}} \approx 12.0$ and changed the apparent statistical significance of paired absolute error differences from $p = 0.0126$ to $p = 0.5492$. Rather than claiming a definitive resolution of all spatial dependencies, this adjustment provides a prudent, conservative check on the evidence base for the current pilot study, demonstrating that the observed performance advantage is not statistically significant under clustered sampling due to both sample size constraints and a potentially modest underlying effect size.

This example highlights a key structural limitation of satellite-based agricultural AI validated against satellite-derived, rather than ground-based, ET references: prediction quality is limited by the spatial and temporal density of the available reference data—in this study, five ECOSTRESS overpasses yielded an effective validation sample of $n_{\text{eff}} \approx 12.0$ from only two adjacent test dates, on which the constrained PIML network did not outperform a simple global-mean predictor. Moreover, this experimental design evaluates short-term temporal transfer at instrumented locations; we cannot claim spatial generalization without field-blocked or leave-one-field-out validation across unobserved sites. Ground-based eddy-covariance validation, which would provide an independent physical measurement rather than a second satellite-derived estimate, remains a priority for future work (Section 4.6). Given these constraints, we present the current results as a proof-of-concept for the double-clip architecture and the design-effect correction methodology,

rather than evidence supporting deployment at regional or continental scales; realizing that broader goal will require continued, spatially independent validation combining multi-year ECOSTRESS and eddy-covariance tower observations, specifically evaluated using field-blocked spatial splits and cross-site transfer learning protocols.

Author Contributions: Conceptualization, U.T. and F.A.; methodology, H.A.; software A.S.; validation H.A., J.A. and H.A.; formal analysis, I.d.l.T.; investigation, J.A.; resources, L.J.H.R.; data curation, P.H.G.; writing—original draft preparation, U.T. and F.A.; writing—review and editing, U.T., F.A., J.A. and A.S.; visualization, I.d.l.T.; supervision, A.S. and F.A.; project administration, A.S.; funding acquisition, L.J.H.R. and P.H.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Fundación Universidad Europea del Atlántico, Calle Isabel Torres 21, 39011 Santander, Spain (VAT: G39764972). The APC was funded by the Fundación Universidad Europea del Atlántico.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author(s).

Acknowledgments: During the preparation of this manuscript, the author(s) used Grammarly 14.1 and ChatGPT-6.1 for language refinement. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Richardson, K.; Steffen, W.; Lucht, W.; Bendtsen, J.; Cornell, S.E.; Donges, J.F.; Drüke, M.; Fetzer, I.; Bala, G.; von Bloh, W.; et al. Earth beyond six of nine planetary boundaries. *Sci. Adv.* **2023**, *9*, eadh2458. [CrossRef] [PubMed]
2. Friedlingstein, P.; O'Sullivan, M.; Jones, M.W.; Andrew, R.M.; Gregor, L.; Hauck, J.; Le Quéré, C.; Luijkx, I.T.; Olsen, A.; Peters, G.P.; et al. Global carbon budget 2022. *Earth Syst. Sci. Data.* **2022**, *14*, 4811–4900. [CrossRef]
3. McDermid, S.; Nocco, M.; Lawston-Parker, P.; Keune, J.; Pokhrel, Y.; Jain, M.; Jägermeyr, J.; Brocca, L.; Massari, C.; Jones, A.D.; et al. Irrigation in the Earth system. *Nat. Rev. Earth Environ.* **2023**, *4*, 435–453. [CrossRef]
4. Friedlingstein, P.; O'Sullivan, M.; Jones, M.W.; Andrew, R.M.; Hauck, J.; Olsen, A.; Peters, G.P.; Peters, W.; Pongratz, J.; Sitch, S.; et al. Global carbon budget 2020. *Earth Syst. Sci. Data.* **2020**, *12*, 3269–3340. [CrossRef]
5. Romanello, M.; Di Napoli, C.; Drummond, P.; Green, C.; Kennard, H.; Lampard, P.; Scamman, D.; Arnell, N.; Ayeb-Karlsson, S.; Ford, L.B.; et al. The 2022 report of the Lancet Countdown on Health and Climate Change: Health at the Mercy of Fossil Fuels. *Lancet* **2022**, *400*, 1619–1654. [CrossRef] [PubMed]
6. Watts, N.; Amann, M.; Arnell, N.; Ayeb-Karlsson, S.; Beagley, J.; Belesova, K.; Boykoff, M.; Byass, P.; Cai, W.; Campbell-Lendrum, D.; et al. The 2020 report of the Lancet Countdown on health and climate change: Responding to converging crises. *Lancet* **2021**, *397*, 129–170. [CrossRef] [PubMed]
7. Wulder, M.A.; Loveland, T.R.; Roy, D.P.; Crawford, C.J.; Masek, J.G.; Woodcock, C.E.; Allen, R.G.; Anderson, M.C.; Belward, A.S.; Cohen, W.B.; et al. Current status of Landsat program, science, and applications. *Remote Sens. Environ.* **2019**, *225*, 127–147. [CrossRef]
8. Huang, S.; Tang, L.; Hupy, J.P.; Wang, Y.; Shao, G. A commentary review on the use of normalized difference vegetation index (NDVI) in the era of popular remote sensing. *J. For. Res.* **2021**, *32*, 1–6. [CrossRef]
9. Xue, J.; Su, B. Significant remote sensing vegetation indices: A review of developments and applications. *J. Sens.* **2017**, *2017*, 1–17. [CrossRef]
10. Lu, B.; Dao, P.D.; Liu, J.; He, Y.; Shang, J. Recent advances of hyperspectral imaging technology and applications in agriculture. *Remote Sens.* **2020**, *12*, 2659. [CrossRef]
11. Tsouros, D.C.; Bibi, S.; Sarigiannidis, P.G. A review on UAV-based applications for precision agriculture. *Information* **2019**, *10*, 349. [CrossRef]
12. Shakhatreh, H.; Sawalmeh, A.H.; Al-Fuqaha, A.; Dou, Z.; Almaita, E.; Khalil, I.; Othman, N.S.; Khreishah, A.; Guizani, M. Unmanned aerial vehicles (UAVs): A survey on civil applications and key research challenges. *IEEE Access.* **2019**, *7*, 48572–48634. [CrossRef]

13. Shen, C.; Appling, A.P.; Gentine, P.; Bandai, T.; Gupta, H.; Tartakovsky, A.; Baity-Jesi, M.; Fenicia, F.; Kifer, D.; Li, L.; et al. Differentiable modeling to unify machine learning and physical models for geosciences. *Nat. Rev. Earth Environ.* **2023**, *4*, 552–567. [CrossRef]
14. Ayaz, M.; Ammad-Uddin, M.; Sharif, Z.; Mansour, A.; Aggoune, E.-H.M. Internet-of-things (IoT)-based smart agriculture: Toward making the fields talk. *IEEE Access.* **2019**, *7*, 129551–129583. [CrossRef]
15. Mowla, M.N.; Mowla, N.; Shah, A.F.M.S.; Rabie, K.M.; Shongwe, T. Internet of things and wireless sensor networks for smart agriculture applications: A survey. *IEEE Access.* **2023**, *11*, 145813–145852. [CrossRef]
16. Lezoche, M.; Hernandez, J.E.; Alemany Díaz, M.D.M.E.; Panetto, H.; Kacprzyk, J. Agri-food 4.0: A survey of the supply chains and technologies for the future agriculture. *Comput. Ind.* **2020**, *117*, 103187. [CrossRef]
17. Shafi, U.; Mumtaz, R.; García-Nieto, J.; Hassan, S.A.; Zaidi, S.A.R.; Iqbal, N. Precision agriculture techniques and practices: From considerations to applications. *Sensors* **2019**, *19*, 3796. [CrossRef] [PubMed]
18. Benos, L.; Tagarakis, A.C.; Dolias, G.; Berruto, R.; Kateris, D.; Bochtis, D. Machine learning in agriculture: A comprehensive updated review. *Sensors* **2021**, *21*, 3758. [CrossRef] [PubMed]
19. Ahmed, A.A.; Sayed, S.; Abdoulhalik, A.; Moutari, S.; Oyedele, L. Applications of machine learning to water resources management: A review of present status and future opportunities. *J. Clean. Prod.* **2024**, *441*, 140715. [CrossRef]
20. Umutoni, L.; Samadi, V. Application of machine learning approaches in supporting irrigation decision-making: A review. *Agric. Water Manag.* **2024**, *294*, 108710. [CrossRef]
21. Addis, A.; Gessesse, A.A. Downscaling MODIS evapotranspiration into finer resolution using a machine learning approach on a small scale, Ribb watershed, Ethiopia. *Environ. Monit. Assess.* **2024**, *196*, 1183. [CrossRef] [PubMed]
22. Zhao, T.; Wang, S.; Ouyang, C.; Chen, M.; Liu, C.; Zhang, J.; Yu, L.; Wang, F.; Xie, Y.; Li, J.; et al. Artificial intelligence for geoscience: Progress, challenges, and perspectives. *Innovation* **2024**, *5*, 100691. [CrossRef] [PubMed]
23. Zhong, L.; Lei, H.; Yang, J. Development of a distributed physics-informed deep learning hydrological model for data-scarce regions. *Water Resour. Res.* **2024**, *60*, e2023WR036333. [CrossRef]
24. Kumar, V.; Kedam, N.; Sharma, K.V.; Mehta, D.J.; Caloiero, T. Advanced machine learning techniques to improve hydrological prediction: A comparative analysis of streamflow prediction models. *Water* **2023**, *15*, 2572. [CrossRef]
25. Slater, L.J.; Arnal, L.; Boucher, M.-A.; Chang, A.Y.-Y.; Moulds, S.; Murphy, C.; Nearing, G.; Shalev, G.; Shen, C.; Speight, L.; et al. Hybrid forecasting: Blending climate predictions with AI models. *Hydrol. Earth Syst. Sci.* **2023**, *27*, 1865–1889. [CrossRef]
26. Frame, J.M.; Kratzert, F.; Raney, A.; Rahman, M.; Salas, F.R.; Nearing, G.S. Post-processing the national water model with long short-term memory networks for streamflow predictions and model diagnostics. *J. Am. Water Resour. Assoc.* **2021**, *57*, 885–905. [CrossRef]
27. Zhi, W.; Appling, A.P.; Golden, H.E.; Podgorski, J.; Li, L. Deep learning for water quality. *Nat. Water.* **2024**, *2*, 228–241. [CrossRef] [PubMed]
28. Feng, D.; Beck, H.; de Bruijn, J.; Sahu, R.K.; Satoh, Y.; Wada, Y.; Liu, J.; Pan, M.; Lawson, K.; Shen, C. Deep dive into hydrologic simulations at global scale: Harnessing the power of deep learning and physics-informed differentiable models (δHBV-globe1.0-hydroDL). *Geosci. Model Dev.* **2024**, *17*, 7181–7198. [CrossRef]
29. Feng, D.; Beck, H.; Lawson, K.; Shen, C. The suitability of differentiable, physics-informed machine learning hydrologic models for ungauged regions and climate change impact assessment. *Hydrol. Earth Syst. Sci.* **2023**, *27*, 2357–2373. [CrossRef]
30. Guo, Z.; Zhang, H.; Martínez-García, E.; Lv, X.; Laudon, H.; Nilsson, M.B.; Peichl, M. Spatio-temporal dynamics and controls of forest-floor evapotranspiration across a managed boreal forest landscape. *Agric. For. Meteorol.* **2025**, *361*, 110316. [CrossRef]
31. Talaviya, T.; Shah, D.; Patel, N.; Yagnik, H.; Shah, M. Implementation of artificial intelligence in agriculture for optimization of irrigation and application of pesticides and herbicides. *Artif. Intell. Agric.* **2020**, *4*, 58–73. [CrossRef]
32. Killeen, P.; Kiringa, I.; Yeap, T.; Branco, P. Corn grain yield prediction using UAV-based high spatiotemporal resolution imagery, machine learning, and spatial cross-validation. *Remote Sens.* **2024**, *16*, 683. [CrossRef]
33. Muñoz-Sabater, J.; Dutra, E.; Agustí-Panareda, A.; Albergel, C.; Arduini, G.; Balsamo, G.; Boussetta, S.; Choulga, M.; Harrigan, S.; Hersbach, H.; et al. ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. *Earth Syst. Sci. Data.* **2021**, *13*, 4349–4383. [CrossRef]
34. Harris, I.; Osborn, T.J.; Jones, P.; Lister, D. Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Sci. Data.* **2020**, *7*, 109. [CrossRef] [PubMed]
35. Abatzoglou, J.T.; Dobrowski, S.Z.; Parks, S.A.; Hegewisch, K.C. Terraclimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. *Sci. Data.* **2018**, *5*, 170191. [CrossRef] [PubMed]
36. Beck, H.E.; Wood, E.F.; Pan, M.; Fisher, C.K.; Miralles, D.G.; van Dijk, A.I.J.M.; McVicar, T.R.; Adler, R.F. Mswep v2 global 3-hourly 0.1° precipitation: Methodology and quantitative assessment. *Bull. Am. Meteorol. Soc.* **2019**, *100*, 473–500. [CrossRef]
37. Karger, D.N.; Conrad, O.; Böhner, J.; Kawohl, T.; Kreft, H.; Soria-Auza, R.W.; Zimmermann, N.E.; Linder, H.P.; Kessler, M. Climatologies at high resolution for the Earth's land surface areas. *Sci. Data.* **2017**, *4*, 170122. [CrossRef] [PubMed]

38. Rolnick, D.; Donti, P.L.; Kaack, L.H.; Kochanski, K.; Lacoste, A.; Sankaran, K.; Ross, A.S.; Milojevic-Dupont, N.; Jaques, N.; Waldman-Brown, A.; et al. Tackling climate change with machine learning. *ACM Comput. Surv.* **2023**, *55*, 1–96. [CrossRef]
39. Hook, S.; Halverson, G. *ECOSTRESS Gridded Evapotranspiration Instantaneous and Daytime L3 Global 70 m V002*; NASA Land Processes Distributed Active Archive Center: Sioux Falls, SD, USA, 2024. [CrossRef]
40. Hengl, T.; Mendes de Jesus, J.; Heuvelink, G.B.M.; Ruiperez Gonzalez, M.; Kilibarda, M.; Blagotić, A.; Shangguan, W.; Wright, M.N.; Geng, X.; Bauer-Marschallinger, B.; et al. SoilGrids250m: Global gridded soil information based on machine learning. *PLoS ONE* **2017**, *12*, e0169748. [CrossRef] [PubMed]
41. Giardina, F.; Gentine, P.; Konings, A.G.; Seneviratne, S.I.; Stocker, B.D. Diagnosing evapotranspiration responses to water deficit across biomes using deep learning. *New Phytol.* **2023**, *240*, 968–983. [CrossRef] [PubMed]
42. dos Santos, R.A.; Mantovani, E.C.; Bufon, V.B.; Fernandes-Filho, E.I. Improving actual evapotranspiration estimates through an integrated remote sensing and cutting-edge machine learning approach. *Comput. Electron. Agric.* **2024**, *225*, 109258. [CrossRef]
43. Zippenfenig, P. *Open-Meteo.com Weather API*, Version 1.4.0; Zenodo: Geneva, Switzerland, 2023. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.