SSRG International Journal of Civil Engineering
ISSN: 2348-8352/ https://doi.org/10.14445/23488352/IJCE-V13I7P110

Volume 13 Issue 7, 161-175, July 2026

Original Article

Comparison of Artificial Neural Networks and Random Forest for Reference Evapotranspiration Estimation and Its Application to the Water Balance of the Pucapuquio Irrigation System, Huancayo, Peru

Yandira Clara Vasquez Martinez[1], Giovanni Joel Vila Romero[2], Erick Jonathan Yañac Terrel[3], Giancarlo Fernando Meza Terbullino[4]

[1,2,3,4]*School of Civil Engineering, Universidad Continental, Huancayo, Perú.*

[4]*Corresponding Author : gmezat@continental.edu.pe*

Received: 04 April 2026 Revised: 04 June 2026 Accepted: 22 June 2026 Published: 29 July 2026

Abstract - *Climate change causes rising temperatures, changes in precipitation and other climatological variables that are related to reference Evapotranspiration (ET_0). The estimation of ET_0 is crucial for irrigation scheduling and water resource management. In this sense, the Penman-Monteith (PM) method is considered the most precise for ET_0 calculation, but the use of the method is often constrained by limited climatological data. In Peru, the PISCOeo_pm dataset provides ET_0 values for the period 1981–2016 that apply the PM method and presents a baseline to train machine learning models that can use limited data to extend ET_0 estimates to recent years and make precise calculations of water demand requirements for irrigation systems. In this research, a random forest model and a multilayer perceptron neural network are developed using PISCOeo_pm data and then used to estimate recent ET_0 values. Daily estimation performance and climatological consistency revealed that the MLP is superior to the RF model. The MLP-estimated ET_0 values for the 2017–2025 period is then used to calculate and update the water demand in the Pucapuquio irrigation system located in the district of Pucará, in Huancayo province, Peru. The results revealed that water demand has increased in recent years and that the previous water demand for the critical month of August, originally 20,769 m³, has increased to 24,150 m³ under current climatological conditions. Potential mitigation strategies are presented that include increasing irrigation application efficiency or reducing cultivated areas by 15% for critical months, until more robust solutions are implemented, such as evaluating supplementary water sources that can potentially satisfy the new demand or the implementation of more efficient irrigation systems.*

Keywords - *Reference evapotranspiration, Highlands, irrigation requirements, ET_0, Water requirement.*

1. Introduction

Evapotranspiration refers to the phenomenon that involves plant transpiration and evaporation from soil and plant surfaces [1]. Furthermore, the Reference evapotranspiration (ET_0) concept permits evaluation of evapotranspiration without considering crop characteristics of a determined region [2].

Climate change causes rising temperatures and shifts in precipitation patterns that influence evapotranspiration trends [3]. As a consequence, ET_0 estimation is essential for irrigation scheduling, water resource management and hydrological studies [4].

The ASCE (American Society of Civil Engineers) standardizes the Penman-Monteith equation for ET_0 estimation [5]. In the same way, the Food and Agriculture Organization [6] recommends the FAO56 Penman-Monteith (FAO56-PM) method to calculate ET_0. The FAO56-PM is considered the most accurate method and requires multiple meteorological inputs, including air temperature, wind speed, humidity, and solar radiation, which, in regions with incomplete or poor-quality meteorological data, have motivated ET_0 alternative estimation methods [7].

In Peru, air temperature data are readily available and reported by the National Service for Meteorology and Hydrology of Peru (SENAMHI) meteorological stations, while the other variables are quite limited or unmeasured [8]. The Hargreaves-Samani (1985) method relies only on air temperature data, disregarding the influence of other variables, and has demonstrated poor performance in ET_0 calculation for Peruvian highlands in comparison to the FAO56-PM method or Machine Learning (ML) approaches

[9]. In response to that, a previous study improved the Hargreaves-Samani equation by calibrating the radiation coefficient for high-altitude conditions [10]. To estimate ET_0 and consider Peru highland conditions, irrigation project technical studies often rely on the Hargreaves (1975) radiation-temperature with altitude and humidity corrections at high-elevation conditions. A historical review and description of the Hargreaves evapotranspiration equation can be found in reference [11].

ML models have been successfully used to estimate ET_0 with limited meteorological data. In Minas Gerais, Brazil, various ML algorithms were compared to estimate ET_0 using limited meteorological data; the ASCE Penman-Monteith was used as a baseline to compare the estimated ET_0 values [12]. Similarly, in India, a comparison between multiple machine learning models to ET_0 estimation was carried out [13]. In addition, the authors found that ML models that used the gridded climate reanalysis product (ERA5) outperformed those relying solely on meteorological stations for longer estimation horizons. This highlights the value of leveraging high-resolution gridded datasets to overcome meteorological limited data. In that sense, SENAMHI developed the grided ET_0 dataset called PISCOeo_pm for the Peruvian territory at a high spatial resolution of approximately 1km [14]; by using spatial grids, the FAO56-PM is applied grid-by-grid, converting the available variables into the exact FAO-56-PM required inputs.

In this research, using the PISCOeo_pm gridded ET_0 database as a baseline, Random Forest (RF) and Multilayer Perceptron (MLP) ML algorithms are compared for ET_0 estimation; the updated ET0 estimations are then used to evaluate the water balance in an irrigation system, which represents a novelty in literature as most presented ML models are often limited only to obtaining ET_0 but an application to irrigation systems is not always considered. The meteorological variables used to train ML models are obtained from the SENAMHI meteorological station located in the Viques district in Huancayo Province, Peru. Since the PISCOeo_pm dataset is only available for the period 1981-2016, the trained ML models are used to extend ET_0 values for the 2017-2025 period using recent meteorological observations as input; this represents a second methodological novelty, as, to the authors' knowledge, no other study in Peru has been published that shows the application of the PISCOeo_pm data on irrigation systems.

The temporal extension captures the effect of recent meteorological variables on ET_0 and represents an improvement over the original technical study of the Pucapuquio irrigation system, which relied on the empirical temperature-based Hargreaves (1975) method with records limited to the 2000-2015 period. The resulting extended ET_0 series is then applied to evaluate and update the water balance of the Pucapuquio irrigation system in Pucará district, Huancayo; this includes a sensitivity analysis of different irrigation efficiencies for different cultivated areas and the impact on the water demand.

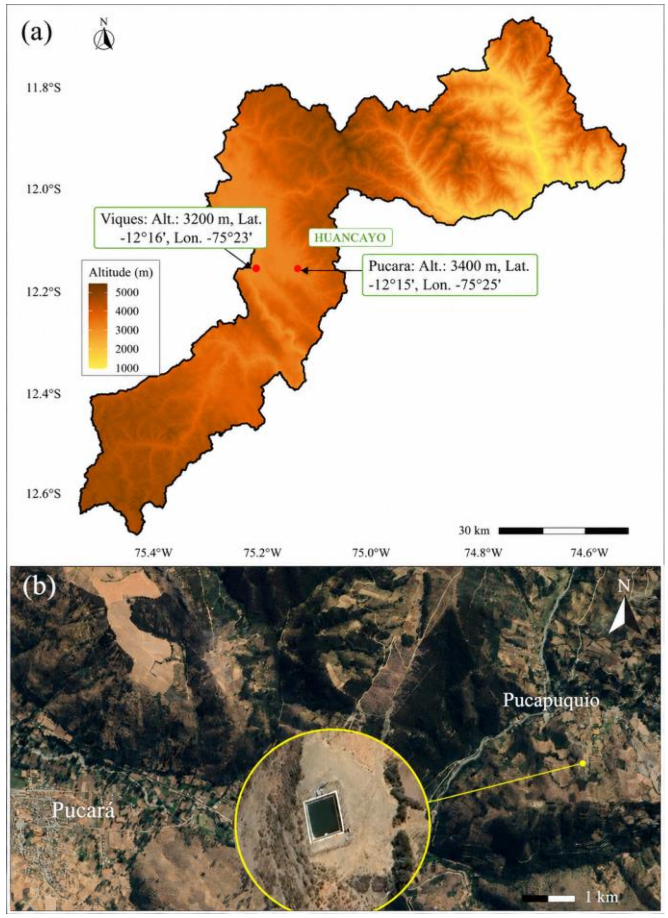


Fig. 1(a) Viques and Pucara location in Huancayo Providence, (b) Reservoir of the Pucapuquio irrigation system.

2. Materials and Methods

2.1. Case Study

Figure 1(a) shows the location of the Pucara district and the Viques district in the province of Huancayo. The SENAMHI meteorological station "Viques "is located at a latitude of 12°09'21.7"S and a longitude of 75°13'41.9"W, and at an elevation of 3186 m. The Viques meteorological station is the nearest station to the Pucapuquio irrigation system, and it provides the meteorological features necessary to calculate the water demand and offer. Figure 1(b) illustrates the reservoir of the Pucapuquio irrigation system.

The Pucapuquio irrigation system is in the community of Pucapuquio in the Pucará district, at approximately 3,500 m above sea level. The system was built through a cooperation agreement between the Peru-Japan Counterpart Fund and the Pucará district municipality. The infrastructure was completed and inaugurated in November 2022 and comprises one intake structure, 3,026 m of conveyance pipeline, a 230 m³ overnight storage reservoir (as shown in Figure 2), 23 pressure-break chambers, and 2,970 m of primary lateral lines with sprinkler units.

Fig. 2 Overnight storage reservoir

2.2. Meteorological Features

As mentioned, to train ML models, meteorological features are obtained from the SENAMHI meteorological station in Viques. These include three meteorological features: daily maximum, minimum temperature, and precipitation. The period 2002-2014 was selected as it corresponds to the range with near-complete daily records for all three variables at Viques station and overlaps with the PISCOeo_pm dataset availability. Specifically, a total of 4,247 daily values from 10-14-2002 to 5-31-2014 were used; the meteorological records are available on the SENAMHI website [8]. From the temperature records, it was calculated daily Delta_T and T_mean, as additional features. The extraterrestrial radiation for daily periods (R_a) and daylight hours or theoretical sunshine (N) were also calculated. In total, 7 meteorological features were used to train the ML models.

Daily meteorological values from the Viques station, 2002-2014, contained sporadic missing values for the three meteorological features. The missing values were imputed using Kalman smoothing on the state-space representation of ARIMA models via the na_kalman function from the imputed TS R package) [15], The model order for each feature is selected by minimizing AIC (Appendix A, Table 1). This approach has been validated for meteorological data imputation [16].

Unlike temperature and precipitation meteorological features, R_a and N are deterministic quantities that depend solely on the station latitude and the day of the year. These were calculated following the FAO-56 methodology. R_a was calculated by using the solar constant, the inverse relative Earth-Sun distance, the solar declination, and the sunset hour angle. N was calculated as a function of the sunset hour angle. The detailed equations can be found in reference [17].

2.2.1. Temperature Trend

The annual mean maximum temperature for the last 9 years (2017-2025) shows an increase over the training period (2002-2014), as shown in the boxplot of Figure 3. The median temperature for the latest years is 21.13°C, while the median temperature for the training period is 20.48°C, supporting the premise that rising temperatures could cause ET_0 variations in recent years.

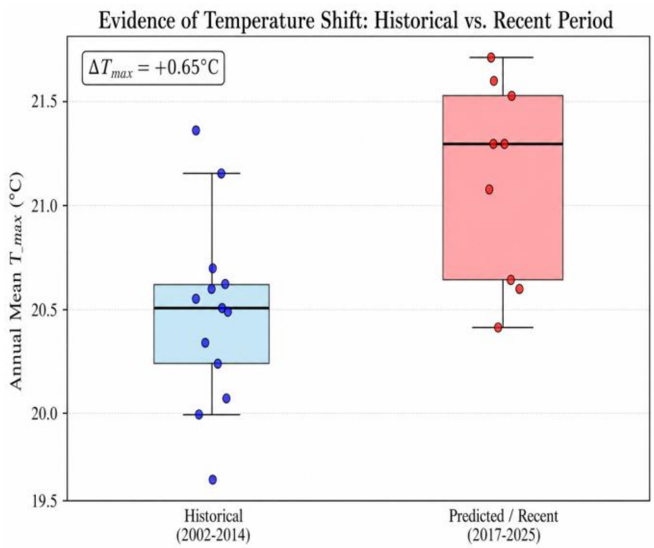


Fig. 1 Comparison of the annual mean maximum temperature between the historical baseline (2002-2014) and the recent period (2017-2025).

2.3. Methodological Framework

Figure 4 depicts the proposed methodological framework. In phase 1, the Viques station meteorological data (extracted from the SENAMHI website) and the PISCOeo_pm data (from the GitHub repository) [18] are prepared. Phase 2 includes developing ML algorithms and comparing the $R^2{}_{train}$ and $R^2{}_{test}$ To ensure robustness of algorithms, the climatological consistency and daily predictive performance are evaluated considering the PISCOeo_pm data (which applies the FAO56-PM method and is physically based).

Subsequently, in phase 3, the best-performing algorithm is used to calculate water demand, and in phase 4, the calculated water demand is compared to the one obtained using the temperature-based Hargreaves (1975) empirical approach; the water supply is based on the Lutz Scholz Model, assuming unchanged water supply conditions for the evaluated period 2017-2025.

Finally, the water balance is obtained for each month, and a sensitivity analysis is carried out for different irrigation efficiencies and cultivated areas. The research is limited to evaluating the new ET_0 methodological approaches recommended by the American Society of Civil Engineers to calculate water demand; the design or enhancement of the irrigation infrastructure is out of the scope of this research.

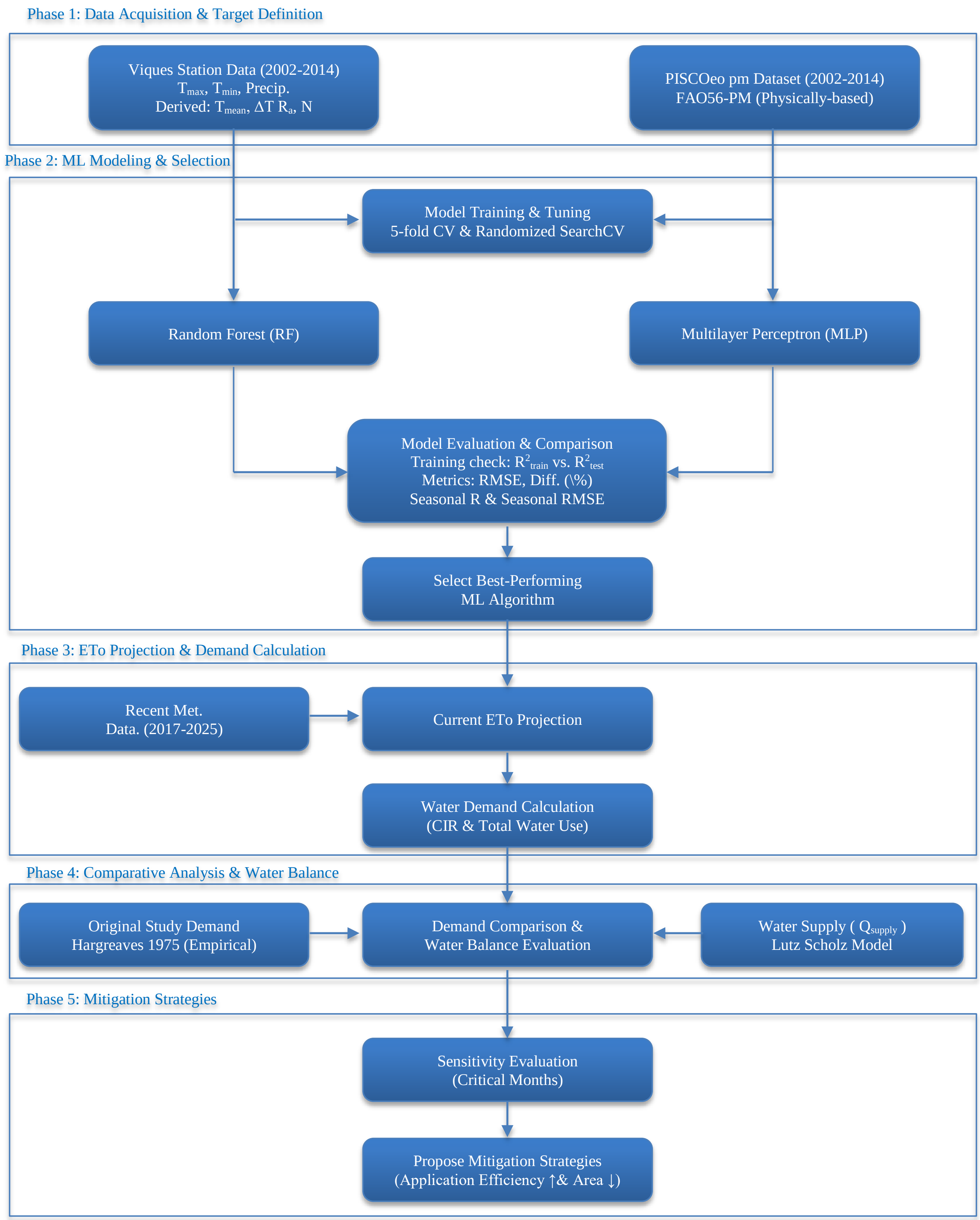


Fig. 2 Methodological framework starting from data acquisition and target definition to water balance evaluation in the Pucapuquio irrigation infrastructure

2.4. Machine Learning Models

2.4.1. Random Forest Model

The RF regressor algorithm is an ML technique based on decision trees, and it is the first model used to estimate ET_o. To ensure generalization and avoid subjective manual tuning, the optimal hyperparameters of the RF model are determined using RandomizedSearchCV, a hyperparameter tuning technique provided by the scikit-learn library in Python. The search process was carried out via a 5-fold Time Series Split cross-validation to respect the chronological order of the meteorological data.

The optimized RF regressor consists of 1250 independent decision trees or estimators, generated using bootstrap sampling. Where each tree independently processes the seven daily meteorological input features to estimate ET_o. The final model's prediction is computed as the average output of all the trees.

To promote diversity among the estimators, the algorithm evaluates a random subset of 50% of the seven features at each splitting node. To avoid overfitting, the maximum depth of each tree was restricted to 8 levels. Furthermore, an internal node is allowed to split if it contains a minimum of 20 daily observations, while final leaf nodes are required to hold at least one sample.

2.4.2. MLP Model

For the MLP model, the RandomizedSearchCV hyperparameter tuning technique combinate with Time Series cross-validation was also utilized. Figure 5 illustrates the optimized MLP architecture, which is defined by an input layer with the seven meteorological features, two hidden layers with 64 and 32 neurons and the predicted ET_o as the output layer. The hidden nodes use the Rectified Linear Unit (ReLU) activation function. Model training uses the Adam optimization algorithm with a constant learning rate of 0.005 and a batch size of 256.

To mitigate overfitting during the training phase, an L2 regularization penalty $\alpha = 0.1$ is applied to constrain the magnitude of the network weights. This constraint prevents any single feature from disproportionately dominates over each other at the ET_o prediction. In addition, an early stop mechanism is used to monitor 10% of the validation subset. This means that if no improvements are observed over 20 consecutive training cycles, the training process automatically halts before reaching the maximum of 1000 iterations.

2.5. Water demand and offer Calculation Procedure

Figure 6 shows the crop schedule in Pucapuquio. For each month, individual crop coefficients $K_{c,i}$ were assigned. This is defined for base crops and rotation crops and their respective cultivated areas. The weight-average crop coefficient $\overline{Kc}$ is calculated as follows:

$$\overline{K_c}(month) = \frac{\sum_{i=1}^{n} K_{c,i}(month) \times A_i}{\sum_{i=1}^{n} A_i} \quad (1)$$

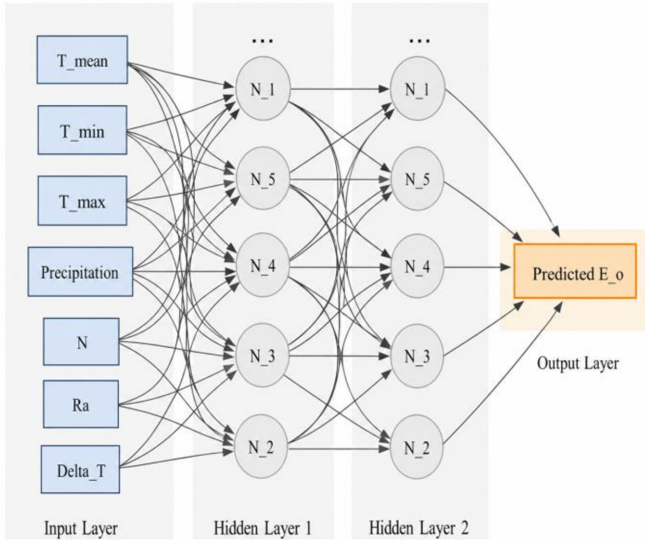


Fig. 3 Conceptual architecture of the optimized MLP neural network showing the seven meteorological input variables and the ET_o output variable

The weight-average crop coefficient is later used to derive total crop evaporation, ET_c, by multiplying by the monthly mean ET_0 as shown in equation 2.

$$ET_c = ET_0\, \overline{K_c} \quad (2)$$

Using the USDA SCS method, the mean monthly precipitation is related to the effective precipitation as shown in equations 3 and 4 [19, 20].

$$P_{eff} = \frac{(125 - 0.2\,P)P}{125} \quad for\ P \leq\ 250\ mm/month \quad (3)$$

$$P_{eff} = 125\ + 0.1P \quad for\ P >\ 250\ mm/month \quad (4)$$

The Crop Irrigation Requirement (CIR) is the water supplied by irrigation that is not provided by precipitation. The CIR values are calculated using equation 5. When $P_{eff} \gg ET_c$, no irrigation is required.

$$CIR\ =\ ET_c\ -\ P_{eff} \quad (5)$$

The Crop Irrigation Requirement (CIR) E_{irr} is the product of conveyance (E_c= 0.95), distribution (E_d=0.95) and application efficiency (E_a), as shown in equation 6. The E_a The value for the pressurized sprinkler system is 0.83. Considering the reported efficiencies in the irrigation project technical study, the overall E_{irr} is 0.75.

$$E_{irr}\ = E_c\, E_d\, E_a \quad (6)$$

The total water use for a month can be obtained from the relation between E_{irr} and CIR [21]:

$$E_{irr} = \frac{CIR}{Total\ Water\ Use} \tag{7}$$

Thereafter, the total irrigation volume in m^3can be obtained by multiplying by the irrigated area (ha) and by the unit conversion factor of 10.

$$V_{demad} = (\,Total\ Water\ Use)(\,A\,)(10) \tag{8}$$

Finally, considering a 16h/day operation for the irrigation system, and days in a month, the flow rate (L/s) is calculated as:

$$Q\ (L/s)\ =\ \frac{V\ (1000)}{d\ (16)(3600)} \tag{9}$$

CROP SCHEDULE IN PUCAPUQUIO

BASE CROPS	NET AREA	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	ROTATION CROP	NET AREA
POTATO	10	1.15	0.75		0.40	0.70	0.95	1.10	0.75		0.41	0.70	1.10	SWEET CORN	5
SWEET CORN	8	0.95	1.10	0.75		0.50	0.90	1.09	1.03	0.51		0.40	0.70	FORAGE OATS	4
WHEAT	3	0.70	1.00	1.20	0.70	0.30	0.40	0.75	1.10	0.90	0.75		0.40	FABA BEAN	2
BARLEY	3	0.70	1.00	1.20	0.60	0.20	0.45	0.70	1.10	1.15	0.75		0.40	GREEN PEA	2
BEAN	2	0.75	1.05	1.15	1.12	0.50							0.55		
FABA BEAN	5	1.10	0.90	0.75	0.30		0.70	0.80	1.00	0.95		0.40	0.75	VEGETABLES	5
PEA	3	0.70	1.10	1.15	0.75	0.40	0.40	0.70	0.95	1.10	0.75		0.45	SWEET CORN	2
TOTAL	**34**	**34**	**34**	**24**	**21**	**20**	**20**	**20**	**20**	**15**	**16**	**23**	**31**		**20**
	$\overline{Kc}$	0.95	0.95	0.95	0.57	0.46	0.72	0.91	0.96	0.87	0.58	0.53	0.81		

Base Crop   Rotation Crop

Fig. 4 Monthly crop schedule in Pucapuquio

To obtain the monthly water supply at Pucapuquio Creek intake, the irrigation project technical study employed the Lutz Scholz deterministic-stochastic rainfall-runoff model. The model was developed for ungauged catchments in Peru highlands, it combinate the deterministic water balance for an average year with a first-order Markov process to produced 28 years series (1988-2015) that incorporates observed precipitation from Viques station from the same years [22]. The synthetic series were used to generate flow duration curves from which $Q_{75\%}$and $Q_{95\%}$Design supply flow at 75% exceedance probability, and ecological flow at 95% exceedance probability, were extracted. The net available supply flow (Q_{supply}) for the irrigation system is obtained as

$$Q_{supply}\ =\ Q_{75\%}\ -\ Q_{95\%}\ -\ \ Q_{third} \tag{10}$$

Where Q_{third} represents existing upstream abstraction by the Puyay Huanca and Pusa Puquio intakes. These are pre-existing creeks diversion structures whose committed water rights must be respected. From Q_{supply}, The monthly supply volume during daytime irrigation hours and the daily nighttime storage are calculated as shown in equations 11 and 12, respectively.

$$V_{16h}\ =\ \frac{16\ (3600)\ d_m\ Q_{supply}}{1000} \tag{11}$$

$$V_{night}\ =\ \frac{8\ (3600)\ Q_{supply}}{1000} \tag{12}$$

Where d_m represents the number of days for a month, and factor 16 is the daytime irrigation hours, while factor 8 is the nighttime hours during which flow accumulates in the reservoir.

The monthly values for Q_{net}, defined as $Q_{75\%}\ -\ Q_{95\%}$, Q_{third}, Q_{supply}, V_{16h}, and V_{night} are presented in Appendix B, Table 1. The annual supply volume amounts to 458,250 m³.

3. Results

3.1. ML Models Comparison

3.1.1. Daily Predictive Performance of ML Models and Residuals Evaluation

The Scatter plots of observed versus predicted ET_0 for MLP in Figure 7(d) show a slope of 0.617, which is closer to unity. In contrast, the RF slope is 0.584 and is shown in Figure 7(a). Regarding distribution and variance of errors, the residual histogram, as shown in Figure 7(b) and Figure 7(e), reveals an approximately symmetric and normal distribution for both models. The mean bias error is low for both MLP is 0.041 mm/day and RF, 0.035 mm/day, respectively.

Finally, the scatter of residuals against predicted values depicted in Figure 7(c) and Figure 7(d) confirms homoscedasticity. The correlation between errors and predictions is r= -0.024 for MLP and r= -0.028 for RF. These values indicate that the models are stable and independent of the magnitude of ET_0.

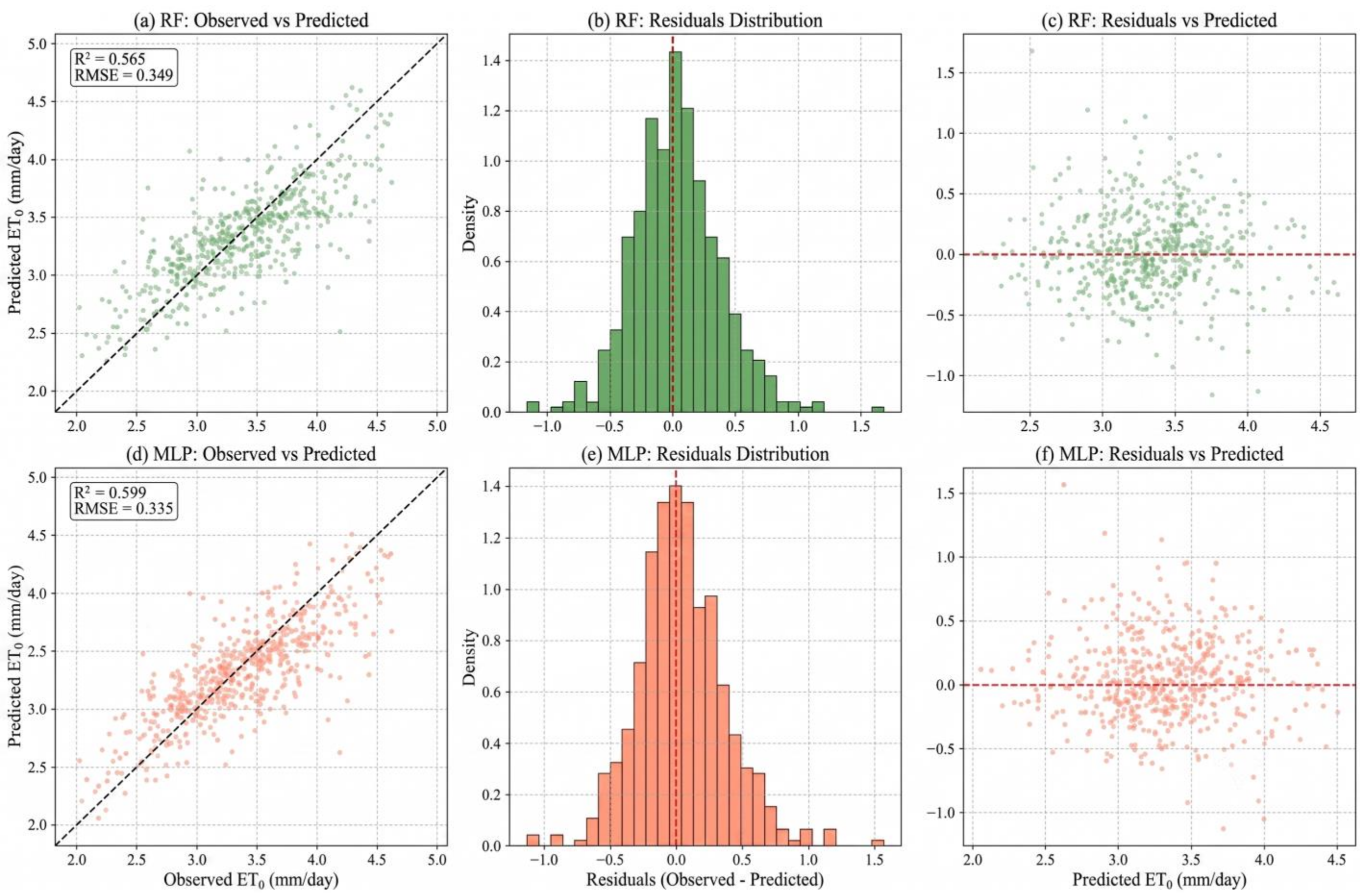


Fig. 5 Daily predictive performance and residual analysis of ML models, (a) Scatter plot of predicted versus observed ET_0 values, where the dashed 1:1 line represents a perfect agreement between predictions and historical records, (b) The probability density histogram of the residuals, (c) The scatter residuals across the range of predicted ET_0 values, (d) Scatter plot of predicted MLP values, (e) Residual distribution histogram for the MLP model, and (f) Residual versus predicted values.

Table 1 shows the daily predictive performance and climatological consistency for both RF and MLP models. Based on R^2 values for the training and testing periods, the RF model tends to overfit, as evidenced by the train-test gap of 0.1769. In contrast, the MLP model presents a train-test gap of 0.059. Similarly, MLP shows lower RMSE metrics for both daily and monthly ET_0 values, 0.335 mm/day and 6.22 mm/month, respectively, whereas the RF model presents 0.349 mm/day and 6.44 mm/month. However, both models show a high seasonal correlation (R) with PISCOeo_pm. The relative difference between the mean annual ET_0 (2002-2014) and the predicted ET_0 for recent years (2017-2025) is marginal: -0.98% for RF and -0.65% for MLP. Although global warming trends might indicate an increase in ET_0, the results imply a stable ET_0. This suggests that the observed increase in temperature may have been offset by changes in other critical meteorological variables.

Table 1. Model evaluation

Metric	RF	MLP
Daily Predictive Performance		
R^2 (Train)	0.742	0.658
R^2 (Test)	0.565	0.599
RMSE (mm/day)	0.349	0.335

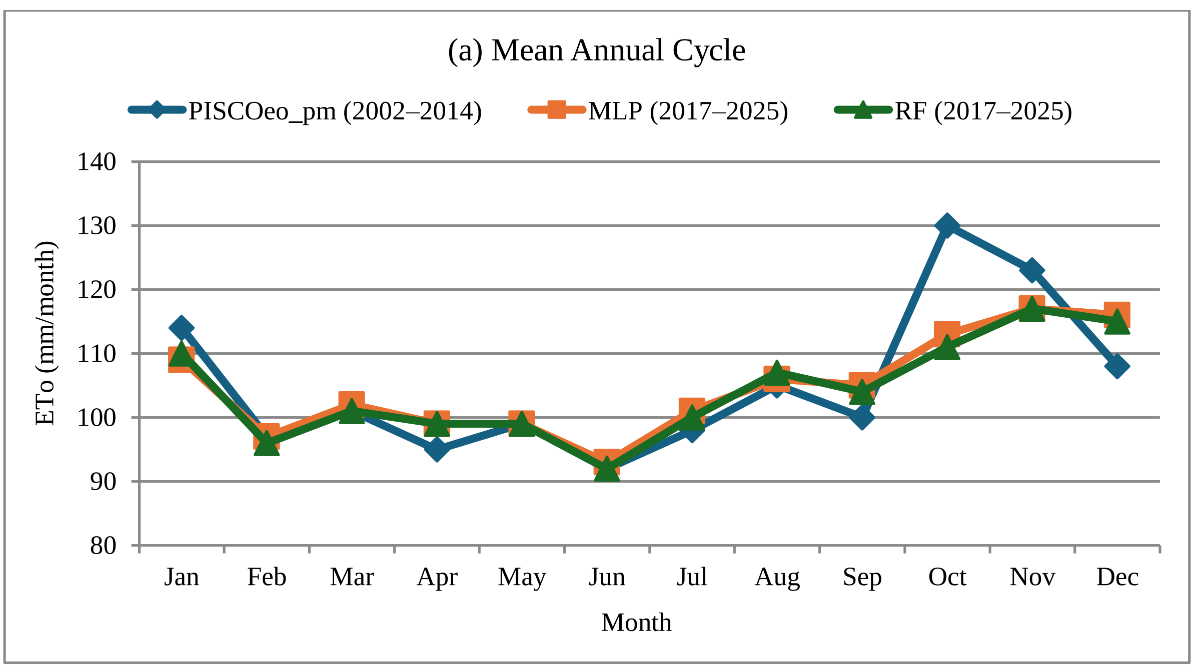
Climatological Consistency		
Relative difference (%)	-0.98%	-0.65%
Seasonal Correlation (R)	0.832	0.846
Seasonal RMSE (mm/month)	6.44	6.22

3.1.2. Comparison of RF and MLP estimated ET₀ values

Table 2 depicts PISCOeo_pm monthly mean ET_o, and the obtained values using RF and MLP models. PISCOeo_pm ET_o Data show a high standard deviation for October and November, 129.7 ± 47.6, and 122.9 ± 37.7 mm/month, respectively. Nevertheless, Figure 8(a) shows that MLP and RF models follow the trend of PISCOeo_pm data. The high standard deviation for October and November in the PISCOeo_pm (2002-2014) is explained by the presence of outliers in October 2003 and October and November 2007, which presented a mean ET_o of 216.06, 210.94, and 234.8 mm/month, respectively. These outliers can be explained as a consequence of the El Niño-Southern Oscillation (ENSO) that impacted those years. On the other hand, the T_{max} values for October 2003 and October and November 2007 are 20.18 °C, 20.56°C and 21.44°C, respectively. These values are not anomalous as they are near the historical average temperature in October and November (2002-2014), 21.21 °C, and 21.41 °C, respectively. Similarly, no observed anomalous precipitation values were observed that could explain these outliers $E.T_o$. This indicates that variables such as air humidity, wind speed, or net solar radiation are likely the cause of the observed outliers in the PISCOeo_pm database. The monthly differences for both models with respect to PISCOeo_pm are shown in Figure 8(b).

Table 2. PISCOeo_pm historical monthly mean ET_o(2002-2014) and predicted RF and MLP ET_o values

Month	PISCOeo_pm (2002-2014)	RF (2017-2025)	MLP (2017-2025)
Jan	113.5 ± 7.6	109.3 ± 4.7	108.4 ± 4.8
Feb	96.8 ± 5.0	95.7 ± 4.5	96.7 ± 4.7
Mar	101.4 ± 3.8	101.5 ± 3.2	101.5 ± 3.3
Apr	95.2 ± 5.2	98.7 ± 3.5	99.4 ± 3.5
May	99.1 ± 4.0	99.1 ± 2.5	99.0 ± 2.5
Jun	92.2 ± 3.8	92.3 ± 4.7	92.8 ± 5.3
Jul	98.4 ± 5.1	100.1 ± 2.7	100.6 ± 3.5
Aug	105.8 ± 4.7	107.0 ± 4.6	106.0 ± 4.8
Sep	100.2 ± 6.3	103.9 ± 3.3	104.6 ± 3.8
Oct	129.7 ± 47.6	110.7 ± 3.0	112.8 ± 3.9
Nov	122.9 ± 37.7	117.1 ± 7.8	116.7 ± 7.9
Dec	107.7 ± 4.8	115.0 ± 6.9	116.1 ± 7.4
Annual	**1262.8**	**1250.5**	**1254.7**
Diff %	**(reference)**	**-0.98%**	**-0.65%**



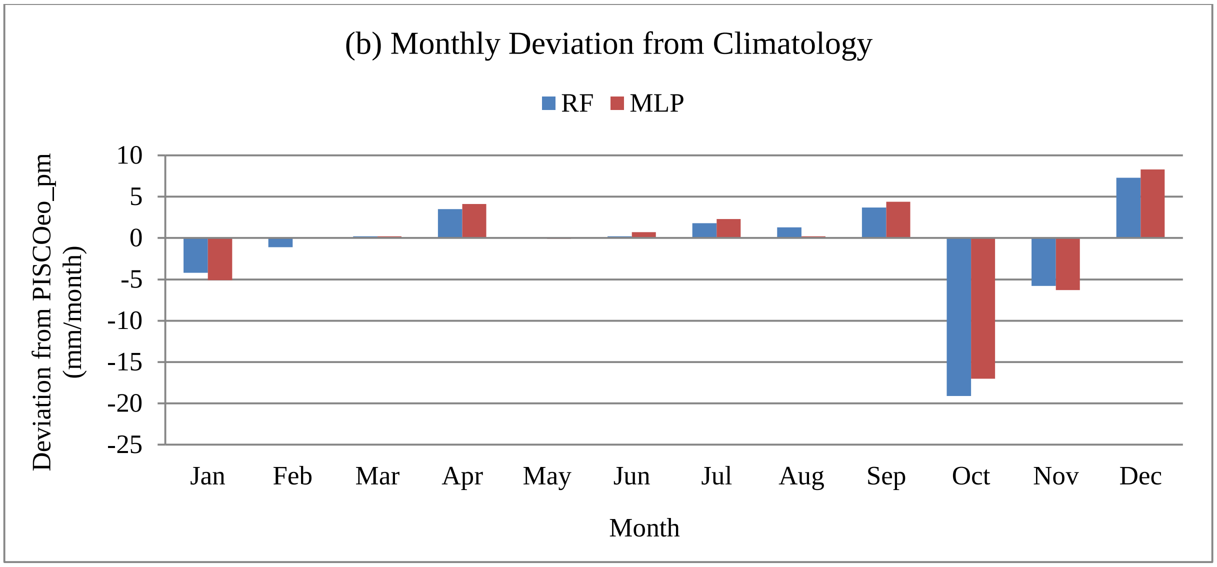


Fig. 6 (a) PISCOeo_pm , RF, and MLP monthly ET_o trend,
(b) Monthly difference between the RF and MLP monthly ET_o values and PISCOeo_pm values.

Figure 8 presents the monthly variability of MLP-predicted ET_o values as boxplots compared against the PISCOeo_pm climatological mean ± 1σ. Median values, Interquartile Range (IQR), 25th percentile (Q25), 75th percentile (Q75), and percentage deviation are shown in Table 3.

In the predicted data set, the following months exhibited outliers ET_o: April, June, and December as shown in Figure 9. In the April months (2017-2025), the lowest mean daily maximum temperature (T_{max}), 20.08 °C, was registered in April 2025; in contrast, the mean daily precipitation was the highest, 2.94 mm/day. A plausible explanation is that high daily precipitation implies increased cloud cover, blocking solar radiation. Consequently, the temperature dropped during that year, leading to a predicted ET_o = 92.761 mm for that month, and becoming an outlier value. The baseline April mean daily. T_{max} and daily precipitation, excluding the April 2025 anomaly, were 21.26°C, and 0.97 mm/day, respectively.

Regarding the June months (2017-2025), June 2017 exhibited the lowest. T_{max}, 18.65 °C, at a low mean daily precipitation value of 0.19 mm/day. Although the precipitation values in June showed low interannual variability, the unusually cold month led to the outlier ET_o = 80.591 mm prediction. The baseline June mean daily. T_{max} and daily precipitation, excluding the 2017 anomaly, were 21.01 °C and 0.19 mm/day.

December 2025 exhibited the highest. T_{max}, 24.11 °C, and an unusually low mean daily precipitation value of 1.11 mm/day. Because of the lower precipitation and the unusually warm conditions, the predicted ET_o = 134.229 mm was identified as an outlier. The boxplot in Figure 10 clearly shows these three ET_o outlier values.

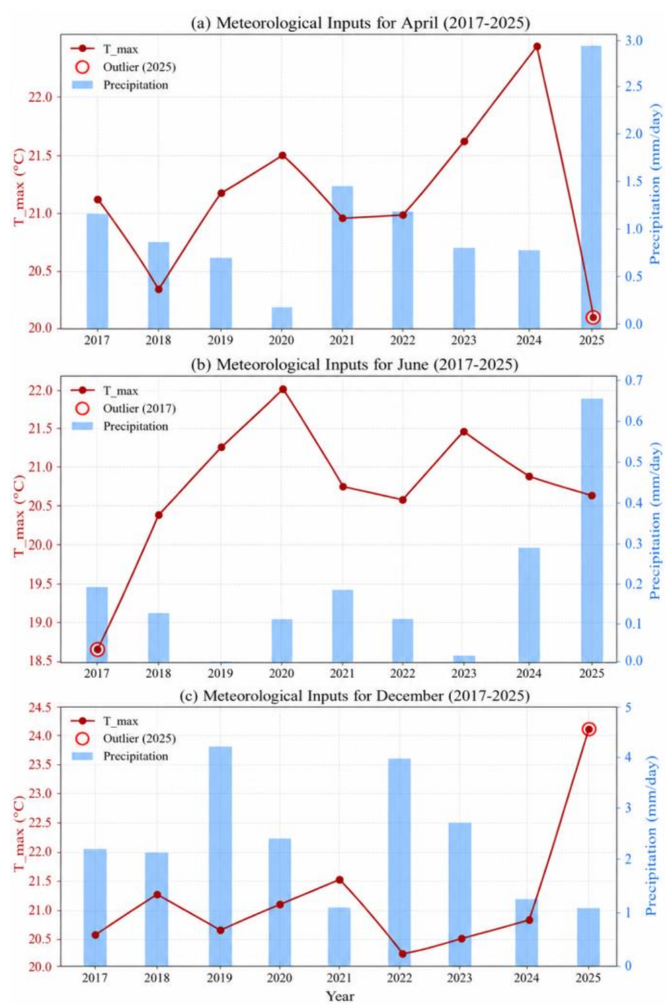


Fig. 7 Anomalous mean daily maximum temperature values that caused outlier values in the predicted MLP dataset (2017-2025). (a) In April 2025, (b) June 2017, and (c) December 2025.

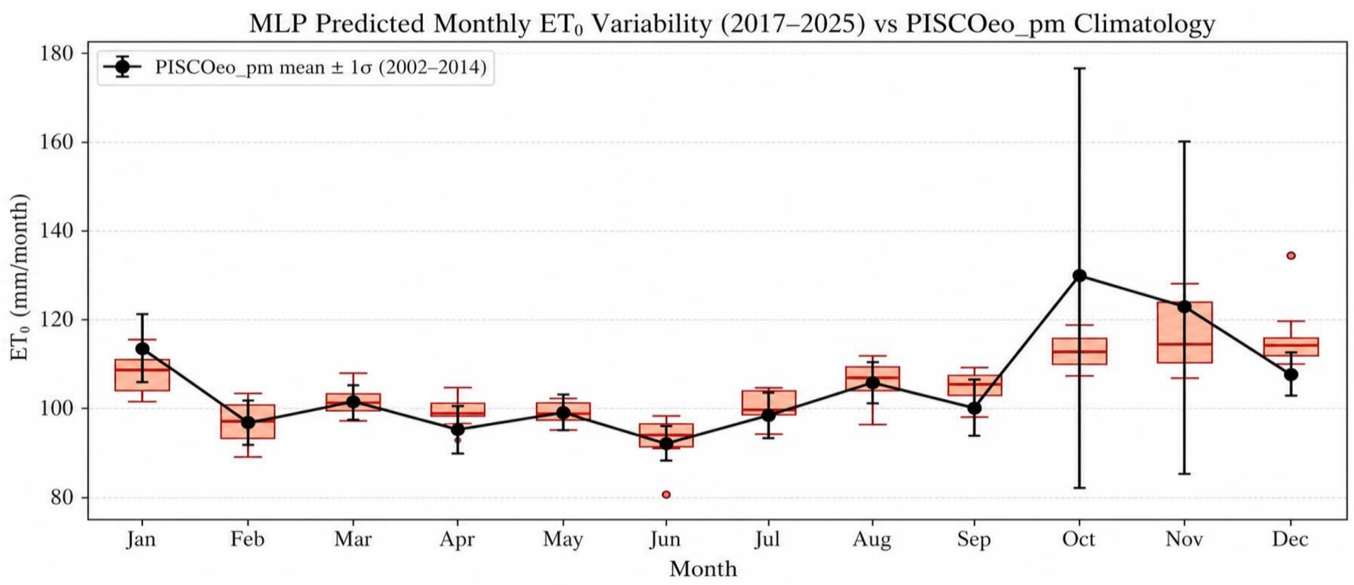


Fig. 8 Boxplot diagram for MLP predicted monthly ET_o values (2017 – 2025) and PISCOeo_pm dataset (2002 – 2014) historical standard deviation

Table 3. Monthly ET_o climatological comparison (mm/month)

Month	PISCOeo_pm Mean ± 1 σ	MLP Median (2017-2025)	MLP IQR [Q25 - Q75]	Deviation (%)	Outliers
Jan	113.5 ± 7.6	108.7	[103.9 - 110.9]	-4.2%	—
Feb	96.8 ± 5.0	97.2	[93.3 - 100.8]	+0.5%	—
Mar	101.4 ± 3.8	102.0	[99.3 - 103.3]	+0.6%	—
Apr	95.2 ± 5.2	98.7	[98.2 - 101.2]	+3.6%	2025
May	99.1 ± 4.0	99.1	[97.4 - 101.1]	-0.0%	—
Jun	92.2 ± 3.8	94.1	[91.3 - 96.8]	+2.1%	2017
Jul	98.4 ± 5.1	99.5	[98.6 - 104.0]	+1.2%	—
Aug	105.8 ± 4.7	107.0	[104.0 - 109.5]	+1.1%	—
Sep	100.2 ± 6.3	105.5	[103.1 - 107.5]	+5.3%	—
Oct	129.7 ± 47.6	112.7	[109.9 - 115.8]	-13.1%	—
Nov	122.9 ± 37.7	114.3	[110.3 - 124.2]	-7.0%	—
Dec	107.7 ± 4.8	114.5	[111.9 - 115.9]	+6.3%	2025

3.2. Water Balance Results

Using the MLP-predicted monthly mean ET_0 For water demand calculation, as explained in section 2.4, the total irrigation volume demand (V_{demand}) was obtained as presented in Appendix C, Table C1. The total irrigation volume for each month is compared to the available monthly supply volume for irrigation. V_{16h}. The resulting monthly and daily balances are shown in Table 4. The balance analysis identifies July, August, and September as deficit months. August represents the most critical case: the total volume demand is 24,150 m³ against an available month supply of only 13,856 m³. As a result, this month, evidence a negative month balance of -10,294 m³ and a daily deficit of -332.1 m³. Furthermore, in August, the overnight reservoir filling capacity only reaches 223.5 m³, yielding a net daily balance of -108.6 m³. This negative balance will result in the reservoir eventually emptying within a few days under sustained demand.

In contrast, the previous irrigation volume demand obtained by using the Hargreaves method was 20,769 m³, which, under the same August volume supply, resulted in a daily balance of -223 m³ and, considering the same filling capacity if 223.5 m³, the net reservoir volume is 0.5 m³.

Figure 11(a) compares the monthly water supply with the irrigation demand calculated using the Hargreaves method and the MLP estimated. ET_0. It is clear that the demand has increased in the critical months of July and August, when the available supply is at its lowest.

Figure 11(b) shows the daily supply-demand balance for each month. For August, the daily deficit shifted from -223 m³ to -332.1 m³, while for July it was originally -85.8 m³ and now is -269.6 m³. However, September now shows a more favorable situation than before, originally -42.5 m³, and now -4.1 m³.

Table 4. Water balance results for each month

Month	V_{demand} (m³/mo)	V_{16h} (m³/mo)	Mo. bal. (m³/mo)	Daily bal. (m³/d)	Night fill (m³/d)	Net res. (m³/d)
Jan	1910	73208	+71298	+2299.9	-	-
Feb	0	98373	+98373	+3513.3	-	-
Mar	4838	50671	+45833	+1478.5	-	-
Apr	6211	31337	+25126	+837.5	-	-
May	8549	21695	+13146	+424.1	-	-
Jun	14839	15403	+564	+18.8	-	-
Jul	22173	13816	-8357	-269.6	222.8	-46.7
Aug	24150	13856	-10294	-332.1	223.5	-108.6
Sep	11504	11380	-124	-4.1	189.7	+185.5
Oct	1331	41486	+40155	+1295.3	-	-
Nov	618	38388	+37770	+1259.0	-	-
Dec	5904	48637	+42733	+1378.5	-	-
Total	**102027**	**458250**	**+356223**			

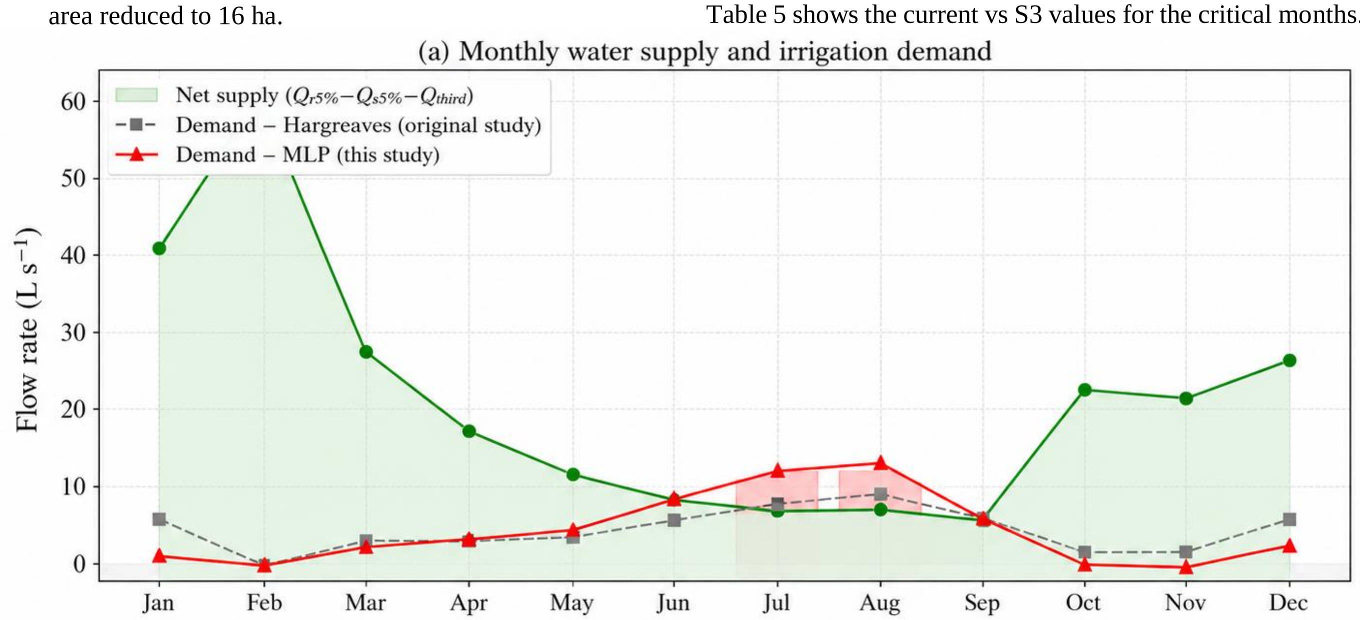
3.3. Scenario Sensitivity Evaluation

The critical months of July and August have a scheduled cultivated area of 20ha. Under the current system's irrigation efficiency $E_{irr} = 0.75$, the only efficiency term with an improvement margin is the application efficiency $E_a = 0.83$ as the other efficiency terms $E_c = 0.95$ and $E_d = 0.95$ would require major pipeline modification, and the improvement margin is almost nonexistent. A modest improvement in the current aspersion efficiency is achievable through pressure regulation at each sprinkler head, regular nozzle maintenance and replacement, and optimal sprinkler overlap adjustment.

Considering a modest application efficiency increase from $E_a = 0.83$ to $E_a = 0.85$, which represents a 2.3% improvement through sprinkler management practice, three scenarios are evaluated:

- Scenario 1: Improved efficiency of $E_a = 0.85$ and the same cultivated area of 20ha.
- Scenario 2: Same efficiency of $E_a = 0.83$ and cultivated area reduced to 16 ha.
- Scenario 3: Improved efficiency of $E_a = 0.85$ and a cultivated area of 17 ha.

Figure 12(a) shows that for the Hargreaves method, the entire parameter space yields positive reservoir balances. In sharp contrast, Figure 12(b) reveals that the MLP-based ET₀ demand presents restricted improvement areas. Under scenario 1, improving only the application efficiency still results in S1 being located in the negative net daily reservoir balance area. Although S2 is in a feasible area, it requires an area reduction from 20 to 16 ha. S3 which includes E_a Improvement and an area reduction to 17 ha is the most equilibrated scenario to mitigate the increasing demand in recent years for the critical months. Specifically, in August the under the S3 configuration, the volume demand shifted from the original 24,150 m³ to 20,069 m³, as a consequence, the net daily reservoir balance changed from -108.6 m³ to 23.1 m³, which shows that the reservoir no longer operates at a deficit. Table 5 shows the current vs S3 values for the critical months.



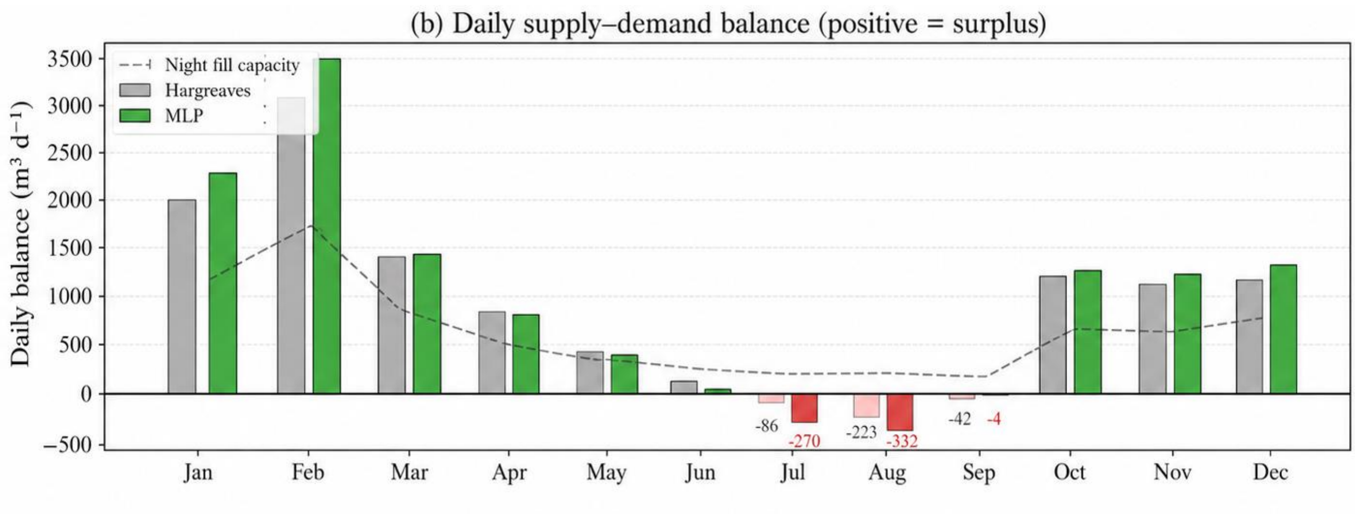


Fig. 11 Water balance for irrigation system, (a) monthly supply volume versus irrigation demand calculated using the Hargreaves method and the MLP-estimated ET₀, (b) daily supply-demand balance for each month under both methods and night fill capacity.

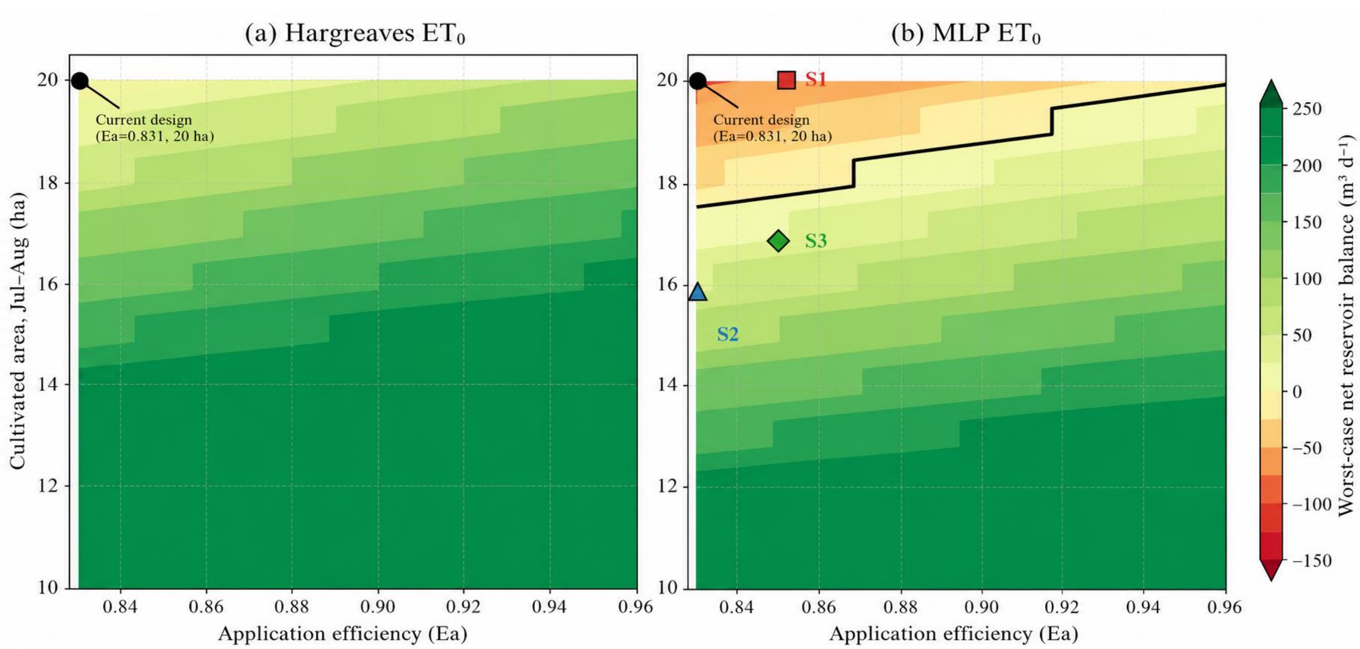


Fig. 12 Sensitivity analysis at an application efficiency and cultivated area for august under, (a) Hargraves, and (b) MLP-based ET₀. Ec=0.95 and Ed=0.95 are held constant.

Table 5. Monthly water balance comparison between the current and design conditions and scenario 3

Month	Scenario	V_demand (m³/mo)	V_supply (m³/mo)	DailyBal (m³/d)	V_night (m³/d)	Net (m³/d)	Status
Jun	Current	14,840	15,403	+18.8	256.7	--	Feasible
	S3	14,509		+29.8		--	Feasible
Jul	Current	22,174	13,816	-269.6	222.8	-46.8	Deficit
	S3	18,427		-148.7		+74.1	Feasible
Aug	Current	24,150	13,856	-332.1	223.5	-108.6	Deficit
	S3	20,069		-200.4		+23.1	Feasible
Sep	Current	11,504	11,380	-4.1	189.7	+185.5	Feasible
	S3	11,247		+4.4		--	Feasible

4. Conclusion

In the presented study, RF and MLP models were developed using seven meteorological features to estimate monthly ET_0 for the period 2017-2025. The PISCOeo_pm ET_0 dataset were used as the target variable for both models. The MLP model showed better performance in daily predictive performance and climatological consistency. For the MLP, the R^2 on test was 0.6, while for the RF was 0.56, meanwhile the relative annual difference was -0.98% for RF and -0.65% for MLP with respect the PISCOeo_pm dataset. These results indicate MLP daily ET_0 estimations are sufficient for water demand calculation, as it is used the mean monthly ET_0 and daily predictive errors tend to cancel out at the monthly scale.

August is identified as the most critical month, as it exhibits the highest total volume demand. The irrigation project technical study originally used the Hargreaves (1975) method and temperature records from 2000-2015, and estimated an August demand. V_{demand} of 20,769 m³. In contrast, by using the MLP and temperature records from 2017-2025 and the other meteorological features, the estimated August demand is 24,150 m³. The difference in the demand values reflects the change climatological conditions over time and the methodological improvement of training the MLP model using the PISCOeo_pm ET_0 dataset, which applies the FAO56-PM method. Under the previous August demand, the reservoir with a night filling capacity of 223.5 m³ was enough to supply the daily demand of -223 m³. However, with the current climatology for the period 2017-2025, the daily demand is -332.1 m³, and under the same filling night capacity, the net reservoir water balance is in deficit. A sensitivity analysis was undertaken to assess adaptation strategies under current climatological conditions.

A potentially viable scenario to mitigate the increase in demand is to reduce cultivated areas from 20ha to 17ha, couple with a 2.3% enhancement in irrigation application efficiency, which is achievable through sprinkler management practices. This results in August demand of 20,069 m³and a net reservoir balance of 23.1 m³, eliminating the deficit. However, this represents only one of several possible adaptation pathways. Other alternatives include evaluating supplementary water sources, harvesting rainwater, more efficient irrigation systems like drip irrigation, or shifting to lower water-demand crops for the critical months. The evaluation of these strategies is beyond the scope of the study. The presented results highlight the impact of climate change in ET_0 as this variable depends on climatological conditions that are evolving over time, causing an increase in ET_0 values and thereby an increasing water demand in the Pucapuquio irrigation system.

References

[1] Kaicun Wang, and Robert E. Dickinson, "A Review of Global Terrestrial Evapotranspiration: Observation, Modeling, Climatology, and Climatic Variability," *Reviews of Geophysics*, vol. 50, no. 2, pp. 1-54, 2012. [CrossRef] [Google Scholar] [Publisher Link]

[2] Richard G. Allen et al., "Crop Evapotranspiration-Guidelines for Computing Crop Water Requirements-FAO Irrigation and Drainage Paper 56," *FAO - Food and Agriculture Organization of the United Nations* 1998. [Google Scholar] [Publisher Link]

[3] Stavroula Dimitriadou, and Konstantinos G. Nikolakopoulos, "Evapotranspiration Trends and Interactions in Light of the Anthropogenic Footprint and the Climate Crisis: A Review," *Hydrology*, vol. 8, no. 4, pp. 1-22, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[4] Ziyuan Gong et al., "A Review of Interactions between Irrigation and Evapotranspiration," *Ecological Indicators*, vol. 169, pp. 1-17, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[5] Richard G. Allen et al., *ASCE's Standardized Reference Evapotranspiration Equation*, Watershed Management and Operations Management, pp. 1-203, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[6] L. S. Pereira et al., *Crop Evapotranspiration, Guidelines for Computing Crop Water Requirements*, Second Edition, Revised 2025, Food and Agriculture Organization of the United Nations, pp. 1-440, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[7] Chul-Gyum Kim et al., "Calibration and Evaluation of Alternative Methods for Reliable Estimation of Reference Evapotranspiration in South Korea," *Water*, vol. 16, no. 17, pp. 1-24, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[8] Hydrometeorological Data, Senamhi, 2026. [Online]. Available: https://www.senamhi.gob.pe/?&p=estaciones

[9] Wilber Laqui et al., "Can Artificial Neural Networks Estimate Potential Evapotranspiration in Peruvian Highlands?," *Modeling Earth Systems and Environment*, vol. 5, no. 4, pp. 1911-1924, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[10] Apolinario Lujano, Miguel Sanchez-Delgado, and Efrain Lujano, "Improvement of Hargreaves–Samani Reference Evapotranspiration Estimates in the Peruvian Altiplano," *Water*, vol. 15, no. 7, pp. 1-16, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[11] George H. Hargreaves, and Richard G. Allen, "History and Evaluation of Hargreaves Evapotranspiration Equation," *Journal of Irrigation and Drainage Engineering*, vol. 129, no. 1, pp. 53-63, 2003. [CrossRef] [Google Scholar] [Publisher Link]

[12] Lucas Borges Ferreira, and Fernando França da Cunha, "New Approach to Estimate Daily Reference Evapotranspiration based on Hourly Temperature and Relative Humidity Using Machine Learning and Deep Learning," *Agricultural Water Management*, vol. 234, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[13] Nehar Mandal, and Kironmala Chanda, "Performance of Machine Learning Algorithms for Multi-Step Ahead Prediction of Reference Evapotranspiration Across Various Agro-Climatic Zones and Cropping Seasons," *Journal of Hydrology*, vol. 620, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[14] Adrian Huerta et al., "PISCOeo_pm, A Reference Evapotranspiration Gridded Database based on FAO Penman-Monteith in Peru," *Scientific Data*, vol. 9, pp. 1-18, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[15] Steffen Moritz, and Thomas Bartz-Beielstein, "imputeTS: Time Series Missing Value Imputation in R," *The R Journal*, vol. 9, no. 1, pp. 207-218, 2017. [Google Scholar] [Publisher Link]

[16] E Afrifa-Yamoah et al., "Missing Data Imputation of High-Resolution Temporal Climate Time Series Data," *Meteorological Applications*, vol. 27, no. 1, pp. 1-18, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[17] Chapter 3 - Meteorological Data, FAO, 2026. [Online]. Available: https://www.fao.org/4/x0490e/x0490e07.htm

[18] A. Huerta, adrHuerta/PISCOeo_pm: PISCOeo_pm, GitHub, 2026. [Online]. Available: https://github.com/adrHuerta/PISCOeo_pm

[19] M. H. Ali, and S. Mubarak, "Effective Rainfall Calculation Methods for Field Crops: An Overview, Analysis and New Formulation," *Asian Research Journal of Agriculture*, vol. 7, no. 1, pp. 1-12, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[20] Tim Hess, "Estimating Green Water Footprints in a Temperate Environment," *Water*, vol. 2, no. 3, pp. 351-362, 2010. [CrossRef] [Google Scholar] [Publisher Link]

[21] Water Resources Program, Program Guidance: Determining Irrigation Efficiency and Consumptive Use, Department of Ecology, 2024. [Online]. Available: https://apps.ecology.wa.gov/publications/summarypages/2011076.html

[22] Sandra del Aguila Rios, Hydrological Modeling of Andean Basins, National University of San Cristóbal de Huamanga, 2024. [Online]. Available: https://isbn.bnp.gob.pe/catalogo.php?mode=detalle&nt=146203

[23] Rob J. Hyndman, Yeasmin Khandakar, "Automatic Time Series Forecasting: The forecast Package for R," *Journal of Statistical Software*, vol. 27, no. 3, pp. 1-22, 2008. [CrossRef] [Google Scholar] [Publisher Link]

Appendix A

The intermittent missing values on the daily records of maximum temperature, minimum temperature and precipitation from the Viques stations are shown in Figure A1. (a-c). The missing values were imputed using the na_kalman function from the imputeTS in R studio [15], which applies Kalman smoothing on the state-space representation of an ARIMA (p,d,q) model. For each meteorological feature the optimal ARIMA order was select using auto. arima [23], which minimized the Akaike Information Criterion (AIC). The selected models and their performance are shown in Table A1. The imputed values are visualized in Figure 1 A1(d-f)

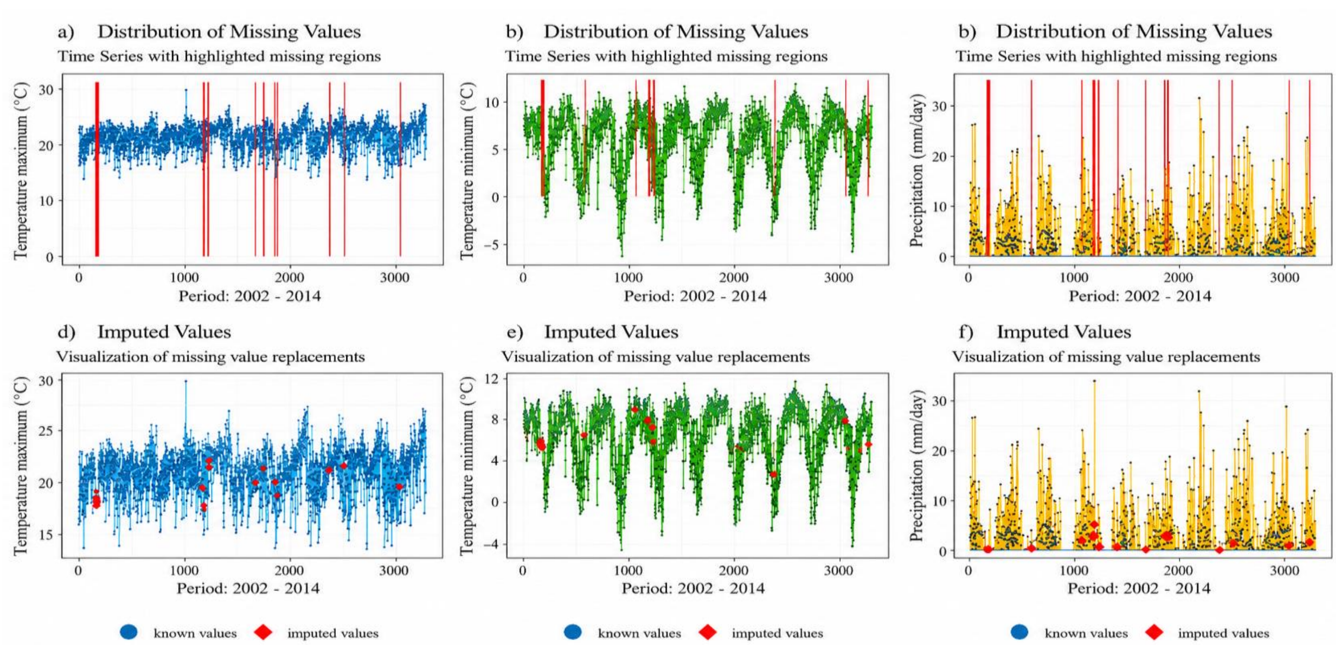


Fig. A1 Identified missing values on the training data set (a-c) and imputed values (d-f).

Table A1. Optimal ARIMA model orders and performance metric for missing values imputation

Variables	AR	D	MA	AIC	RMSE
d)Temp. max	3	0	4	12696.98	1.77 °C
e) Temp. min	2	0	1	13076.02	1.8 °C
f) Precipitation	3	0	1	17505.02	3.59 mm

Appendix B

Table B1. Calculated available water supply values

Month	Q_NET	Q_THIRD	Q_SUPPLY	V_16h	V_night
	(L/s)	(L/s)	(L/s)	(m^3/mo)	(m^3/d)
Jan	47.0	6.00	41.00	73208	1180.8
Feb	61.0	0.00	61.00	98373	1756.7
Mar	31.9	3.52	28.38	50671	817.3
Apr	21.3	3.17	18.13	31337	522.3
May	16.3	4.15	12.15	21695	349.9
Jun	15.5	6.59	8.91	15403	256.7
Jul	15.0	7.26	7.74	13816	222.8
Aug	16.2	8.44	7.76	13856	223.5
Sep	13.4	6.81	6.59	11380	189.7
Oct	24.9	1.67	23.23	41486	669.1
Nov	24.3	2.08	22.22	38388	639.8
Dec	33.6	6.36	27.24	48637	784.5
				458250	

Appendix C

Table C1. Calculated the monthly water demand for irrigation

Month	ET_o	$\overline{K_c}$	ETc	P	P_{eff}	CIR	Area	Net Vol.	Flow
	(mm)		(mm)	(mm)	(mm)	(mm)	(ha)	(m^3)	(L/s)
Jan	108.365	0.95	102.95	122.9	98.73	4.21	34	1910	1.1
Feb	96.747	0.95	91.91	130.5	103.25	-	34	0	0.0
Mar	101.517	0.95	96.44	96.1	81.32	15.12	24	4838	2.7
Apr	99.365	0.57	56.64	36.6	34.46	22.18	21	6211	3.6
May	99.034	0.46	45.56	13.8	13.50	32.06	20	8549	4.8
Jun	92.831	0.72	66.84	11.4	11.19	55.65	20	14839	8.6
Jul	100.587	0.91	91.53	8.5	8.38	83.15	20	22173	12.4
Aug	105.995	0.96	101.76	11.4	11.19	90.56	20	24150	13.5
Sep	104.599	0.87	91.00	35.5	33.48	57.52	15	11504	6.7
Oct	112.809	0.58	65.43	66.2	59.19	6.24	16	1331	0.7
Nov	116.669	0.53	61.83	67.0	59.82	2.02	23	618	0.4
Dec	116.143	0.81	94.08	93.9	79.79	14.28	31	5904	3.3
Total	**1254.66**		**966.0**	**693.8**	**594.3**			**102028**	