

Estimating Monthly Reference Evapotranspiration from Limited Meteorological Data: Machine Learning versus Locally Calibrated Empirical Equations in a Humid Temperate Climate

Somayeh Mirshekari¹✉, Maryam Safavi

¹ Agriculture Institute, Research Institute of Zabol, Zabol, Iran

² Ph.D, Department of Water Engineering, University of Birjand, Birjand, Iran

Corresponding Author, email: Smirshekari@uoz.ac.ir

Received: 07 September 2025

Accepted: 09 October 2025

Available online: 11 December 2025

Abstract: The FAO-56 Penman–Monteith (PM) equation is the standard for reference evapotranspiration (ET₀), but it requires radiation, humidity and wind data that are often missing. This study used 24 years (1976–1999) of daily weather records from the Wageningen (Haarweg) station in the Netherlands, aggregated to 284 complete months. It compared six empirical equations (Hargreaves–Samani, Thornthwaite, Oudin, Makkink, Priestley–Taylor and Turc), in original and locally calibrated form, with five data-driven models (multiple linear regression, support vector regression (SVR), Gaussian process regression (GPR), XGBoost and a multilayer perceptron) for six input combinations. All models were evaluated by six-fold blocked cross-validation using 4-year blocks. Monthly PM-ET₀ averaged 631 mm yr⁻¹, peaking at 104 mm in July. Annual T_{max}, T_{min}, actual vapour pressure and wind speed increased significantly, whereas the increase in ET₀ (+4.0 mm yr⁻¹) was marginal ($p = 0.051$). Sensitivity analysis showed that ET₀ was most responsive to vapour pressure (−13.4% for +10%) and maximum temperature (+8.1% for +1 °C). The original empirical equations had RMSEs of 9.1–13.8 mm month⁻¹, with Hargreaves–Samani overestimating by 15% and Makkink underestimating by 15%. Local calibration reduced the errors to 4.4 mm month⁻¹ (Makkink) and 6.1 mm month⁻¹ (Hargreaves–Samani). With temperature and extraterrestrial radiation only, GPR reached RMSE = 5.3 mm month⁻¹ (NSE = 0.977), better than the calibrated Hargreaves–Samani equation. Adding vapour pressure reduced the error to 3.1 mm month⁻¹, and the full input set gave 1.4 mm month⁻¹ (NSE = 0.998). GPR was the best model for every input combination, while XGBoost was the weakest non-linear model. In this humid climate, humidity information is at least as valuable as radiation for data-driven ET₀ estimation, and a calibrated Makkink equation remains a strong radiation-based alternative.

Keywords: Reference evapotranspiration, FAO-56 Penman–Monteith, Gaussian process regression, Hargreaves–Samani, Makkink, Sensitivity analysis

1. Introduction

Reference evapotranspiration (ET₀) is the basis for estimating crop water requirements, scheduling irrigation and closing catchment water balances. The FAO-56 Penman–Monteith (PM) equation (Allen et al., 1998) is the internationally recommended standard, but it requires air temperature, solar radiation, humidity and wind speed. Many stations, especially in developing regions and in historical archives, record only temperature and precipitation, and radiation and humidity series are often incomplete. Simpler equations based on temperature (Thornthwaite, 1948; Hargreaves and Samani, 1985; Oudin et al., 2005) or radiation (Makkink, 1957; Priestley and Taylor, 1972; Turc, 1961) have therefore remained in wide use. Their accuracy varies considerably between climates, and local calibration is usually required (Tabari et al., 2013).

Machine learning (ML) models have been proposed as a way to approximate PM-ET₀ from reduced input sets. Kumar et al. (2002) showed that artificial neural networks can reproduce PM estimates. Later studies compared heuristic regression methods (Kisi, 2016), random forests and generalized regression neural networks (Feng et al., 2017), and tree-based ensembles, support vector machines and extreme learning machines

under limited data in different climates of China (Fan et al., 2018). Comparative reviews generally find that ML models outperform uncalibrated empirical equations (Granata, 2019; Ferreira et al., 2019). Some of these comparisons, however, validate ML models with random data splits while applying empirical equations without calibration, which can favour the ML models. Gaussian process regression (GPR; Rasmussen and Williams, 2006), which is well suited to small samples and smooth relationships, has also received less attention.

At the monthly scale, the resolution used in many water-balance and planning studies, sample sizes are small (a few hundred values per station), which increases the risk of overfitting. Humid climates such as north-western Europe pose a further challenge: ET₀ is dominated by radiation, and the Makkink equation is used operationally in the Netherlands (de Bruin and Lablans, 1998). Changes in temperature, humidity, radiation and wind also affect ET₀ trends, as discussed for the widely reported decline in pan evaporation (Roderick and Farquhar, 2002; McVicar et al., 2012). Sensitivity analysis helps to identify which variables matter most (McCuen, 1974; Gong et al., 2006).

This study used a 24-year daily record from Wageningen, the Netherlands, to (i) characterise the monthly climatology, trends and sensitivity of PM-ET₀; (ii) evaluate six empirical equations before and after local calibration; (iii) compare five data-driven models across six input combinations with the same blocked cross-validation; and (iv) interpret the models by permutation importance.

2. Materials and Methods

2.1. Study Site and Data

The data came from the Wageningen (Haarweg) agro-meteorological station (51.97°N, 5.67°E, 7 m a.s.l.) in the central Netherlands (Figure 1). The climate is humid temperate oceanic. The daily records for 1976–1999 contain global radiation (kJ m⁻² d⁻¹), minimum and maximum air temperature (°C), early-morning vapour pressure (kPa), mean wind speed at 2 m (m s⁻¹) and precipitation. They are distributed as reference weather files (source: Natuur- en Weerkunde, Wageningen) with the Python Crop Simulation Environment (PCSE), which implements the WOFOST crop model.

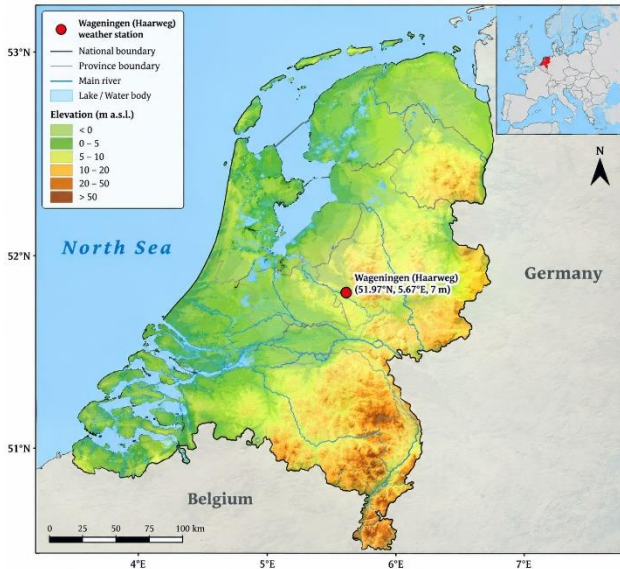


Figure 1. Location of the Wageningen (Haarweg) weather station

Quality control identified 8 spurious records in 1989 (all fields equal to 1 or 3, duplicating valid days), which were removed. It also found 4 missing vapour-pressure values and 5 missing wind-speed values (coded -99), which were filled by linear interpolation over gaps of up to three days. Records for 1 September to 31 December 1991 (122 days) were missing, so these four months were excluded. The final dataset had 284 complete months. The monthly mean variables used as model inputs were T_{max}, T_{min}, solar radiation R_s (MJ m⁻² d⁻¹), actual vapour pressure e_a (kPa), wind speed u₂ (m s⁻¹) and extraterrestrial radiation R_a (MJ m⁻² d⁻¹). Relative humidity was computed as e_a/e_s for description only.

2.2. Reference Evapotranspiration Equations

Daily ET₀ was computed with the FAO-56 PM equation (Allen et al., 1998) and summed to monthly totals, which served as the target:

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

where Δ is the slope of the saturation vapour pressure curve, $R_{\{n\}}$ is net radiation computed from $R_{\{s\}}$, $T_{\{max\}}$, $T_{\{min\}}$ and $e_{\{a\}}$ with an albedo of 0.23, $G = 0$ at the daily scale, γ is the psychrometric constant for 7 m elevation, and $e_{\{s\}}$ is the mean saturation vapour pressure. The implementation reproduced FAO-56 Example 18 ($R_a = 41.09$ MJ m⁻² d⁻¹, $R_n = 13.28$ MJ m⁻² d⁻¹, $ET_0 = 3.88$ vs. 3.9 mm d⁻¹). Six empirical equations were evaluated (Table 1). Five were computed daily and summed, and Thornthwaite was computed from monthly mean temperature with the heat index of the 1976–1999 climatology:

$$ET_{HS} = 0.0023 \cdot 0.408 R_a (T + 17.8) \sqrt{T_{max} - T_{min}} \quad (2)$$

$$ET_{MK} = 0.65 \frac{\Delta}{\Delta + \gamma} \frac{R_s}{\lambda} \quad (3)$$

Table 1. Empirical ET₀ equations evaluated

Method	Inputs	Formulation / reference
Hargreaves–Samani (HS)	T, T _{max} – T _{min} , R _a	Eq. (2); Hargreaves and Samani (1985)
Thornthwaite (TH)	Monthly T, day length	Heat-index formula; Thornthwaite (1948)
Oudin (OU)	T, R _a	$R_a/\lambda \cdot (T + 5)/100$; Oudin et al. (2005)
Makkink (MK)	T, R _s	Eq. (3); Makkink (1957), de Bruin and Lablans (1998)
Priestley–Taylor (PT)	T, R _n	$1.26 \cdot \Delta / (\Delta + \gamma) \cdot R_n / \lambda$; Priestley and Taylor (1972)
Turc (TU)	T, R _s	$0.013 \cdot T / (T + 15) \cdot (23.8846 \cdot R_s + 50)$, RH > 50%; Turc (1961)

$\lambda = 2.45$ MJ kg⁻¹. Priestley–Taylor uses the same net radiation as PM.

Each empirical estimate E was also calibrated locally by linear regression, $ET_0 = a \cdot E + b$, using the same cross-validation folds as the ML models.

2.3. Data-Driven Models and Input Combinations

Six input combinations were defined: C1 (T_{max}, T_{min}, R_a), C2 (C1 + R_s), C3 (C1 + e_a), C4 (C1 + u₂), C5 (C1 + R_s + e_a) and C6 (all inputs). C1 corresponds to a temperature-only station, because R_a depends only on latitude and date. Five models were compared: multiple linear regression (MLR); SVR with a radial basis function kernel (Vapnik, 1995); GPR with an anisotropic squared-exponential kernel plus white noise; XGBoost (Chen and Guestrin, 2016); and a multilayer perceptron (MLP) trained with L-BFGS. The GPR predictive mean is (Rasmussen and Williams, 2006):

$$\bar{f}_* = \mathbf{K}_*^T (\mathbf{K} + \sigma_n^2 \mathbf{I})^{-1} \mathbf{y} \quad (4)$$

where $\mathbf{K}$ is the kernel matrix of the training inputs, $\mathbf{K}_*$ holds the covariances between the new input and the training inputs, and σ_n^2 is the noise variance. The kernel was

$$k(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \exp \left[-\frac{1}{2} \sum_d \frac{(x_d - x'_d)^2}{\ell_d^2} \right] + \sigma_n^2 \delta(\mathbf{x}, \mathbf{x}') \quad (5)$$

GPR hyperparameters (σ_f , length scales ℓ_d , σ_n) were estimated by maximising the marginal likelihood with three restarts. For SVR ($C \in \{1, 10, 100, 1000\}$, $\gamma \in \{0.01, 0.05, 0.1, 0.5\}$, $\varepsilon \in \{0.01, 0.05, 0.1\}$), XGBoost (200 or 500 trees, learning rate 0.03 or 0.1, depth 2 or 4, minimum child weight 1 or 5) and MLP (hidden layers (5), (10) or (10, 5); $\alpha \in \{10^{-4}, 10^{-2}, 1\}$; tanh or ReLU), grid search was carried out with an inner five-fold cross-validation grouped by year. Inputs were standardised.

2.4. Validation, Trends and Sensitivity

All models were evaluated by six-fold blocked cross-validation. The folds were the 4-year blocks 1976–1979, 1980–1983, ..., 1996–1999, and each block was predicted by models trained on the other five. The out-of-fold predictions for all 284 months were compared with PM-ET0 using R^2 , RMSE, MAE, NSE, KGE, PBIAS and RSR = RMSE/SD (Moriiasi et al., 2007):

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (O_i - P_i)^2}, \quad PBIAS = 100 \frac{\sum_i (P_i - O_i)}{\sum_i O_i} \quad (6)$$

Trends in annual values for the 23 complete years were tested with the Mann–Kendall test, using the Hamed and Rao (1998) correction where serial correlation was present, and Sen's slope. Sensitivity was assessed by perturbing each daily input (Tmax and Tmin by $+1^\circ\text{C}$; Rs, ea and u2 by $+10\%$) and recalculating PM-ET0:

$$\delta ET_0 = 100 \frac{ET_0(x+\Delta x) - ET_0(x)}{ET_0(x)} \quad (7)$$

Permutation importance (30 repetitions per fold) was computed for GPR, SVR and XGBoost with the full input set (C6).

3. Results

3.1. Climatology, Trends and Sensitivity

Monthly PM-ET0 ranged from 5.9 to 139.1 mm month⁻¹ (mean 52.9; Table 2). It followed the radiation cycle, from 13 mm in December to 103 mm in July (Figure 2). Solar radiation was the variable most strongly correlated with monthly ET0 ($r = 0.975$), followed by Tmax (0.896). Relative humidity ($r = -0.872$) and wind speed (-0.523) were negatively correlated, because both are highest in winter. The mean relative humidity of 77% confirms the humid character of the site.

Table 2. Descriptive statistics of monthly variables (284 months, 1976–1999)

Variable	Min	Max	Mean	SD	CV	Skew	r with ET0
Tmax (°C)	-1.23	27.24	13.56	6.56	0.48	-0.10	0.896
Tmin (°C)	-7.65	14.58	5.33	4.83	0.91	-0.19	0.788
Rs (MJ m ⁻² d ⁻¹)	1.27	22.71	9.52	6.01	0.63	0.20	0.975
ea (kPa)	0.33	1.76	1.01	0.32	0.32	0.22	0.758
u2 (m s ⁻¹)	1.46	5.19	2.93	0.67	0.23	0.90	-0.523
RH (%)	52.47	92.75	77.39	8.00	0.10	-0.48	-0.872
ET0-PM (mm month ⁻¹)	5.95	139.15	52.92	35.10	0.66	0.46	1.000

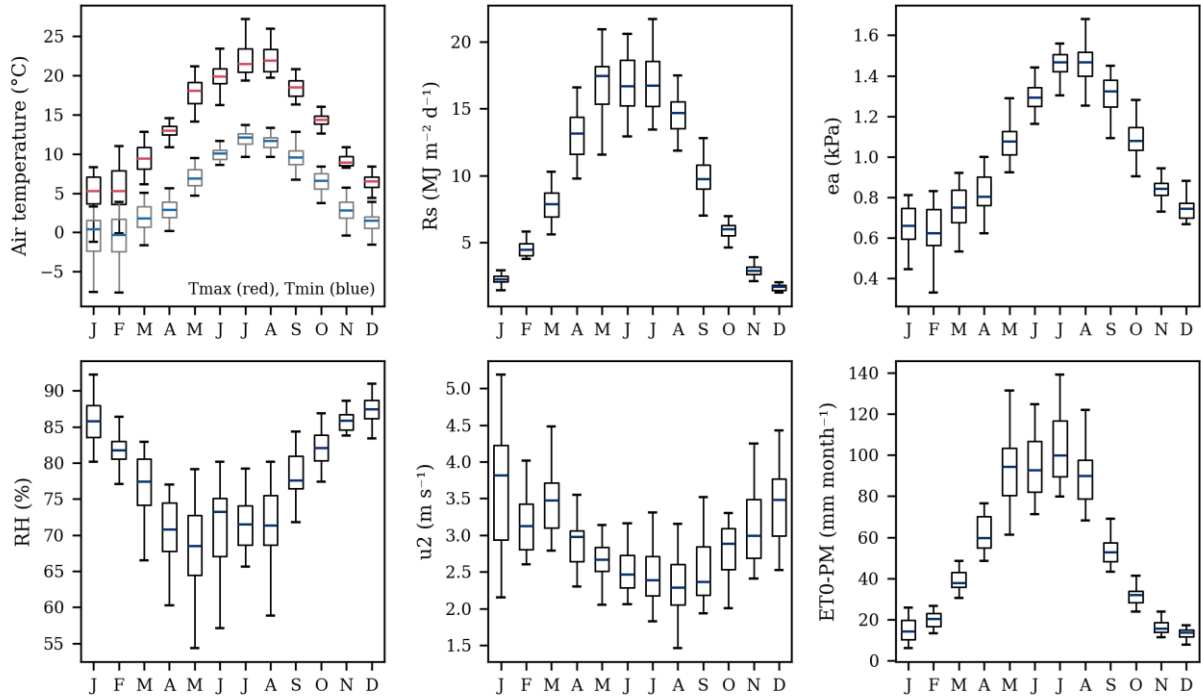


Figure 2. Monthly distributions of temperature, solar radiation, vapour pressure, relative humidity, wind speed and PM-ET0

Annual PM-ET0 averaged 631 mm (SD 54 mm), with a minimum of 549 mm in 1981 and a maximum of 730 mm in 1976 (Figure 3; Table 3). Annual Tmax (+0.63 °C per decade), Tmin (+0.51 °C per decade), vapour pressure and wind speed increased significantly. The ET0 increase (+4.0 mm yr⁻¹) was marginally non-significant ($p = 0.051$), as was the increase in solar radiation ($p = 0.064$). The rise in vapour pressure partly offset the effect of warming on the vapour pressure deficit, so relative humidity did not change significantly.

Table 3. Trends in annual means (annual total for ET0), 23 complete years

Variable	Mean	Z	p	Sen slope (yr ⁻¹)
ET0-PM (mm)	630.89	1.95	0.051	+4.0201
Tmax (°C)	13.54	3.11	0.002	+0.0633
Tmin (°C)	5.34	2.85	0.004	+0.0508
Rs (MJ m ⁻² d ⁻¹)	9.45	1.85	0.064	+0.0483
ea (kPa)	1.01	2.80	0.005	+0.0034
u2 (m s ⁻¹)	2.92	2.01	0.045	+0.0142
RH (%)	77.51	-0.74	0.460	-0.0661

Hamed–Rao modified Mann–Kendall test; the variance correction was effective only for Tmax and Tmin (lag-1 autocorrelation 0.43 and 0.35).

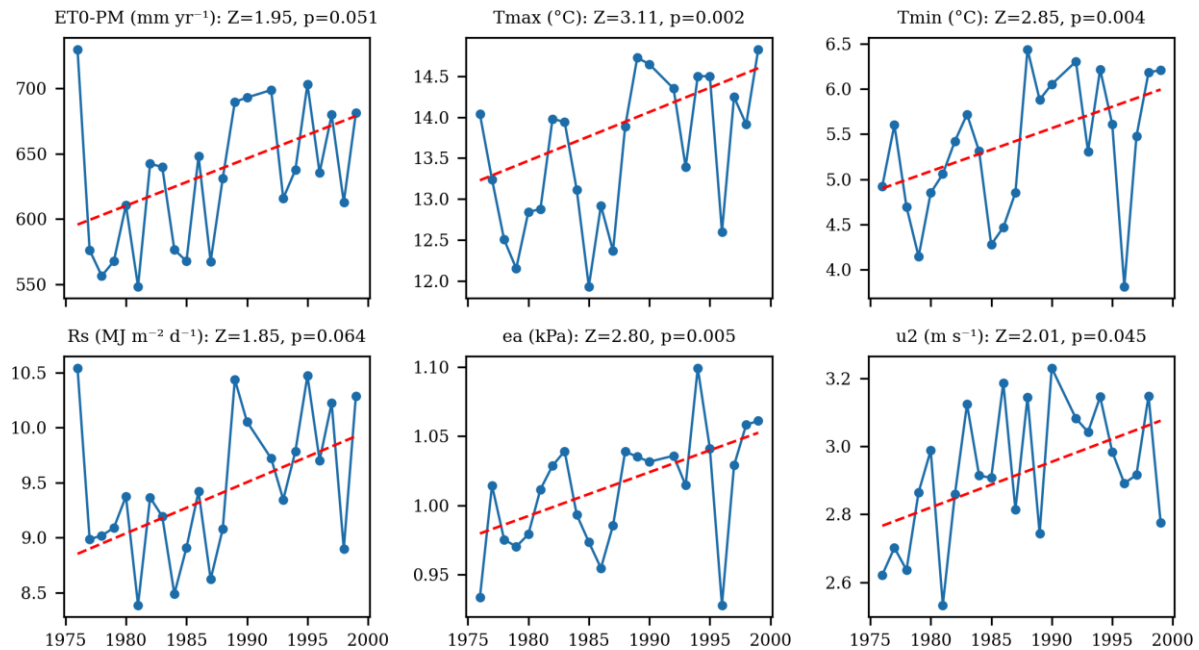


Figure 3. Annual PM-ET0 and meteorological variables with Theil–Sen trend lines

The sensitivity analysis (Figure 4) showed that a 10% increase in vapour pressure reduced annual ET0 by 13.4%, and a 1 °C increase in Tmax raised it by 8.1%. The corresponding changes were 5.2% for Tmin, 3.8% for Rs and 1.9% for wind speed. The relative sensitivities were strongly seasonal. In December, +10% ea reduced ET0 by 51% and +1 °C Tmax increased it by 28%, whereas in July the changes were -8% and +5%. A 10% increase in Rs raised summer ET0 by about 5% but slightly reduced winter ET0 (-3.2% in December), because the higher Rs/Rso ratio increases net longwave loss when shortwave input is small.

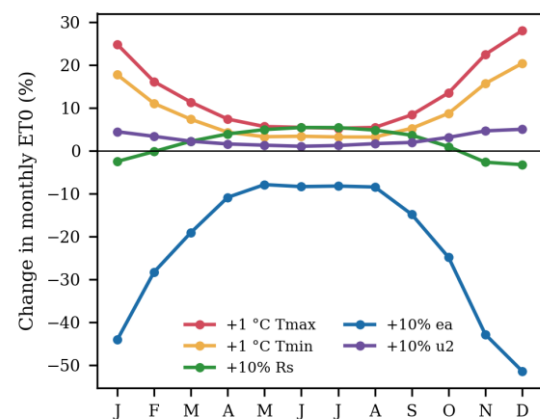


Figure 4. Mean relative change in monthly PM-ET0 for perturbations of each input

3.2. Performance of Empirical Equations

In their original form, the empirical equations had RMSEs of 9.1–13.8 mm month⁻¹ (Table 4; Figure 5). Hargreaves–Samani overestimated ET0 by 14.8%. The radiation-based Makkink (-15.2%), Turc (-13.8%) and Priestley–Taylor (-8.8%) equations underestimated it, although they were very highly correlated with PM (R^2

= 0.973–0.985). Thornthwaite was nearly unbiased (+0.2%) but had the lowest correlation ($R^2 = 0.877$) because it ignores radiation. After local calibration, the Makkink equation was the best empirical method (RMSE = 4.35 mm month⁻¹, NSE = 0.985; $ET_0 \approx 1.07 \cdot MK + 4.8$), followed by Priestley–Taylor (5.31), Turc (5.82) and Hargreaves–Samani (6.05). Calibration improved Thornthwaite and Oudin only slightly (12.34 and 9.13).

Table 4. Performance of the empirical equations before and after local calibration (blocked cross-validation)

Method	Form	R ²	RMS E	MA E	NS E	KG E	PBIAS (%)
Hargreaves–Samani	Original	0.971	13.84	10.57	0.844	0.697	+14.8
	Calibrated	0.970	6.05	4.37	0.970	0.979	-0.0
Thornthwaite	Original	0.877	13.77	11.40	0.845	0.868	+0.2
	Calibrated	0.876	12.34	9.80	0.876	0.910	-0.0
Oudin	Original	0.934	10.29	8.13	0.914	0.899	-6.0
	Calibrated	0.932	9.13	6.55	0.932	0.951	-0.0
Makkink	Original	0.985	9.45	8.08	0.927	0.830	-15.2
	Calibrated	0.985	4.35	3.31	0.985	0.989	+0.0
Priestley–Taylor	Original	0.977	9.08	7.77	0.933	0.823	-8.8
	Calibrated	0.977	5.31	3.97	0.977	0.983	+0.0
Turc	Original	0.973	9.35	7.69	0.929	0.861	-13.8
	Calibrated	0.972	5.82	4.44	0.972	0.980	-0.0

RMSE and MAE in mm month⁻¹.

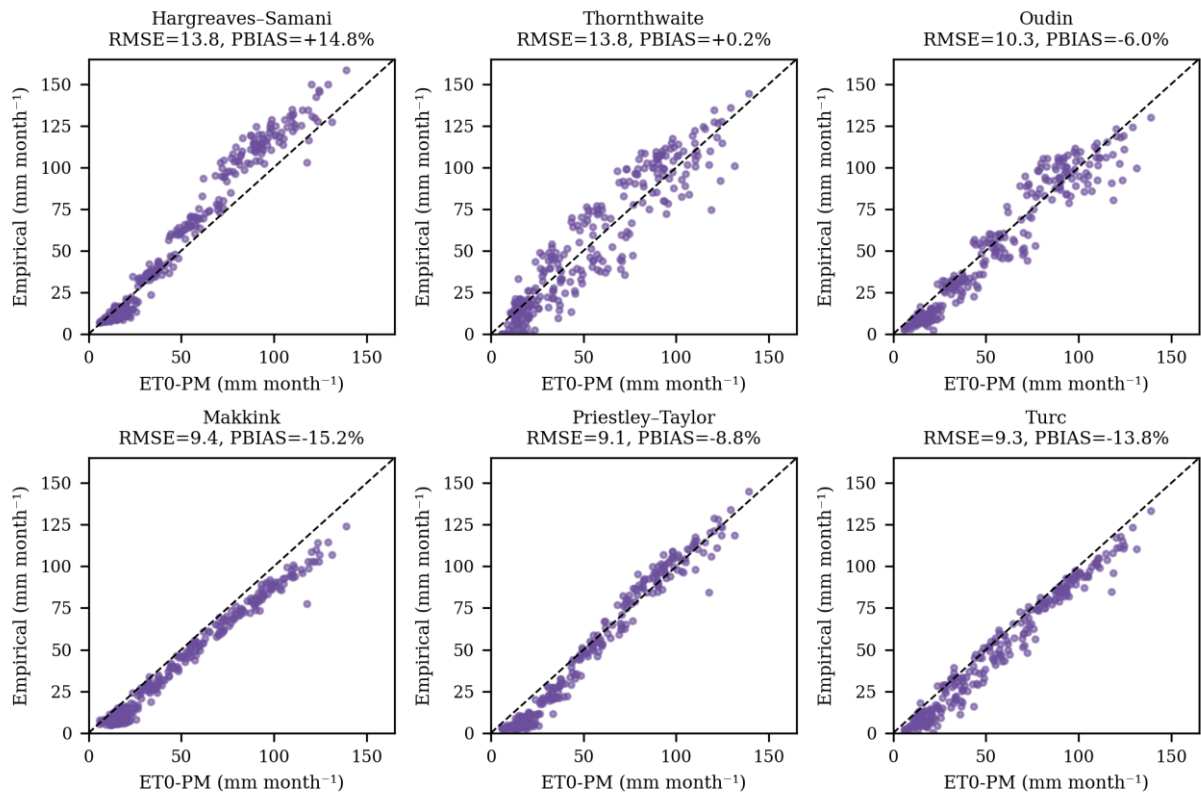


Figure 5. Original empirical ET_0 estimates versus PM- ET_0 (dashed: 1:1 line)

3.3. Performance of Data-Driven Models

Table 5. Cross-validated performance of the data-driven models for six input combinations

Combination	Model	Train NSE	R ²	RMSE	MAE	NSE	KGE	PBIAS (%)
C1	MLR	0.933	0.930	9.26	7.35	0.930	0.950	-0.01
	SVR	0.979	0.976	5.42	3.77	0.976	0.976	-0.91
	GPR	0.980	0.977	5.28	3.74	0.977	0.984	-0.03
	XGBoost	0.990	0.964	6.61	4.80	0.964	0.975	-0.01
C2	MLP	0.981	0.975	5.56	3.98	0.975	0.984	-0.07
	MLR	0.975	0.973	5.76	4.40	0.973	0.981	-0.05
	SVR	0.988	0.986	4.11	2.95	0.986	0.989	-0.57
	GPR	0.989	0.987	4.06	2.96	0.987	0.991	+0.02
	XGBoost	0.995	0.984	4.49	3.34	0.984	0.980	-0.68

C3	MLP	0.989	0.986	4.17	3.05	0.986	0.991	-0.02
	MLR	0.937	0.936	8.89	7.21	0.936	0.950	+0.09
	SVR	0.994	0.991	3.37	2.49	0.991	0.994	-0.15
	GPR	0.995	0.992	3.06	2.24	0.992	0.996	+0.09
C4	XGBoost	0.998	0.983	4.56	3.48	0.983	0.986	-0.44
	MLP	0.994	0.991	3.36	2.43	0.991	0.995	+0.05
	MLR	0.937	0.936	8.85	7.12	0.936	0.952	+0.11
	SVR	0.986	0.984	4.46	3.14	0.984	0.984	-0.55
C5	GPR	0.988	0.985	4.24	3.02	0.985	0.990	-0.10
	XGBoost	0.998	0.971	5.98	4.31	0.971	0.980	-0.20
	MLP	0.987	0.981	4.78	3.43	0.981	0.988	-0.35
	MLR	0.975	0.973	5.73	4.41	0.973	0.981	-0.07
C6	SVR	0.997	0.996	2.25	1.75	0.996	0.997	-0.09
	GPR	0.997	0.997	2.04	1.58	0.997	0.998	-0.00
	XGBoost	0.999	0.988	3.89	2.95	0.988	0.984	-0.71
	MLP	0.997	0.996	2.19	1.68	0.996	0.998	+0.06
C6	MLR	0.977	0.974	5.63	4.24	0.974	0.982	-0.04
	SVR	0.999	0.998	1.56	1.22	0.998	0.997	-0.29
	GPR	0.999	0.998	1.42	1.11	0.998	0.999	+0.01
	XGBoost	0.999	0.990	3.58	2.64	0.990	0.982	-0.73
C6	MLP	0.999	0.997	1.96	1.49	0.997	0.993	+0.25

C1: Tmax, Tmin, Ra; C2: C1 + Rs; C3: C1 + ea; C4: C1 + u2; C5: C1 + Rs + ea; C6: all. RMSE and MAE in mm month⁻¹. Bold: best model per combination.

GPR had the lowest RMSE for every input combination (Table 5; Figure 6). With temperature and Ra only (C1), GPR reached RMSE = 5.28 mm month⁻¹ (NSE = 0.977), close to SVR (5.42) and MLP (5.56). All three were better than the calibrated Hargreaves–Samani equation (6.05) and much better than MLR (9.26). Adding one variable to C1 reduced the GPR error to 4.06 (Rs), 3.06 (ea) and 4.24 mm month⁻¹ (u2). Vapour pressure was therefore the most informative single

addition. This contrasts with the linear models, for which Rs was far more useful than ea (MLR: 5.76 vs. 8.89). Combining Rs and ea (C5) gave 2.04 mm month⁻¹, and the full set (C6) gave 1.42 mm month⁻¹ (NSE = 0.9984, KGE = 0.999). With radiation available (C2), GPR, SVR and MLP were only slightly better than the calibrated Makkink equation (4.35), and XGBoost was worse (4.49).

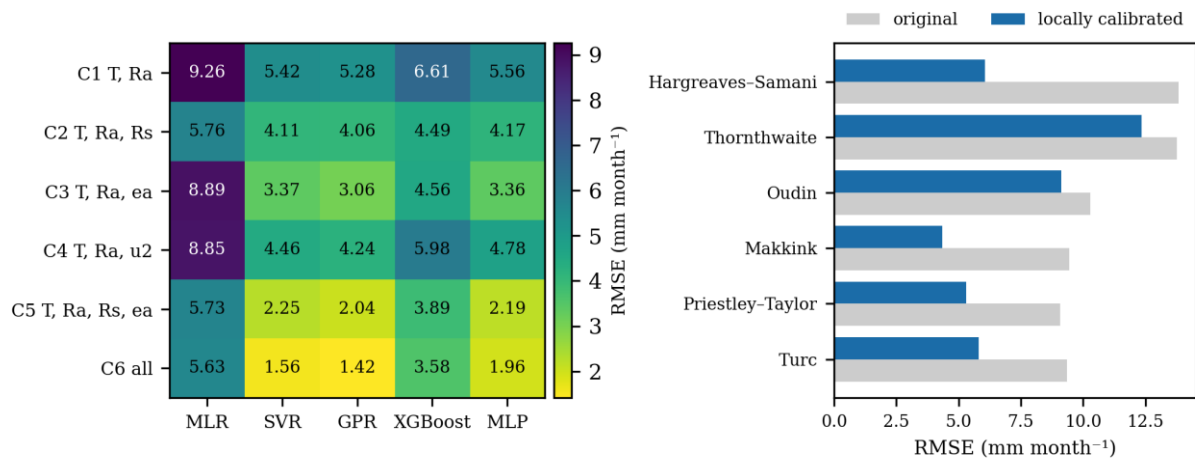


Figure 6. Cross-validated RMSE of the data-driven models for each input combination (left) and of the empirical equations before and after calibration (right)

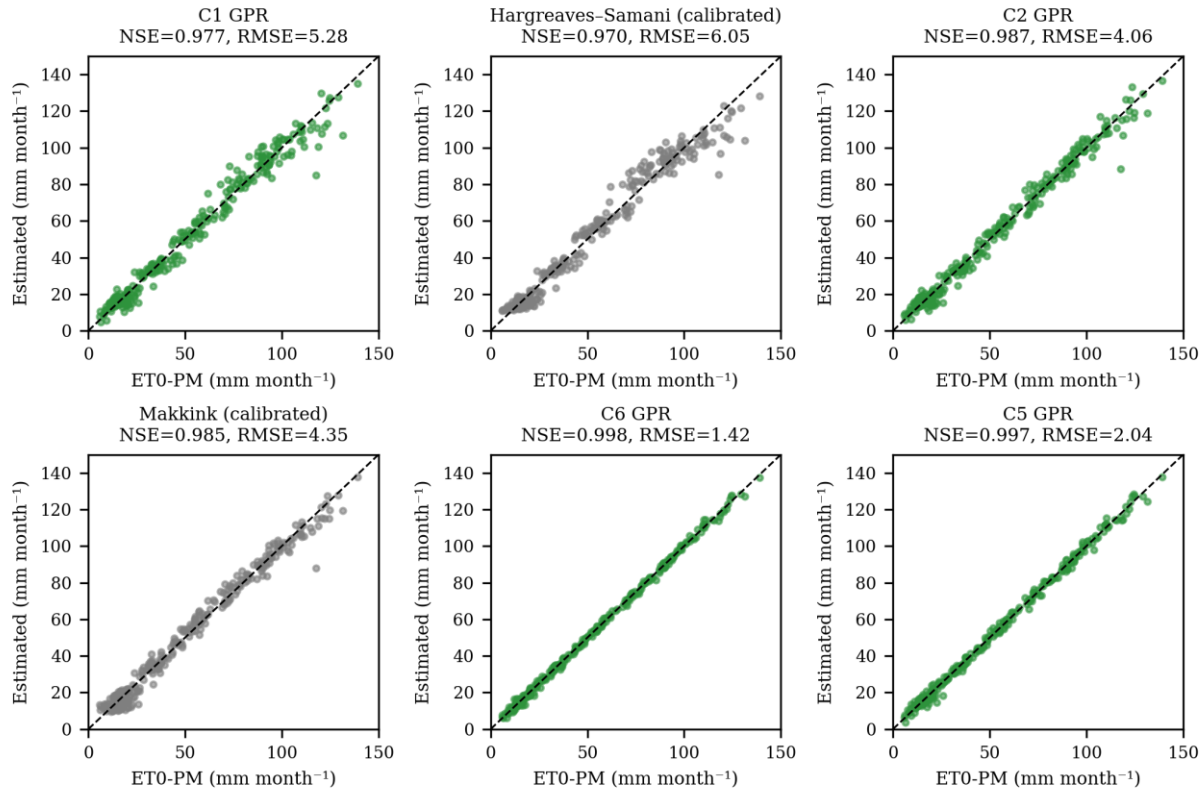


Figure 7. Cross-validated estimates of the best data-driven models (C1, C2, C5, C6) and of the calibrated Hargreaves–Samani and Makkink equations

XGBoost was the weakest non-linear model in every combination (RMSE 3.58–6.61). It had the largest gap between training and validation fit (training NSE ≥ 0.990) and systematically underestimated ET0 (PBIAS down to -0.7%). Its piecewise-constant predictions are poorly suited to the smooth, nearly deterministic relationship between the inputs and PM-ET0. All other models were practically unbiased ($|PBIAS| < 1\%$). Scatter plots (Figure 7) show that the remaining GPR errors in C1 and C2 were largest in the high-ET0 summer months.

The monthly error pattern (Figure 8) confirms this. For the temperature-based C1 GPR and the calibrated Hargreaves–Samani equation, RMSE peaked in May (8.7 and 10.5 mm) and August (9.5 and 9.7 mm), when cloudiness, and hence radiation, varies most for a given temperature. Radiation-based models (C2 GPR, calibrated Makkink) reduced the May error to 4.6–5.6 mm, but their August errors remained about 7 mm.

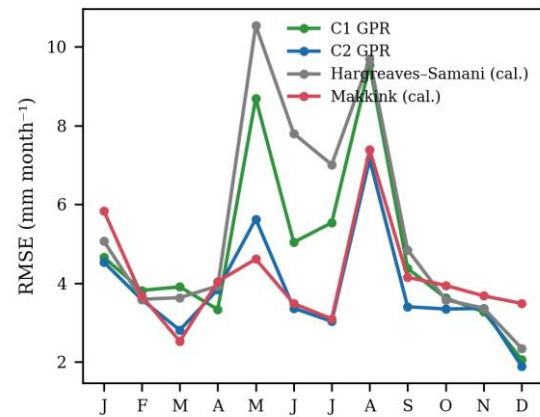


Figure 8. RMSE by calendar month for selected models

3.4. Input Importance

Permutation importance for C6 (Figure 9) differed between model types. For GPR and SVR, Tmax was the most influential input (31% and 32%), followed by ea (26% and 21%), while Ra, Tmin and Rs each contributed 13–16%. For XGBoost, Rs dominated (52%). Wind speed contributed only 1.5–2.0% in all models. The smooth kernel models thus exploited the vapour pressure deficit, i.e. the joint information in Tmax, Tmin and ea, which also explains why C3 outperformed C2.

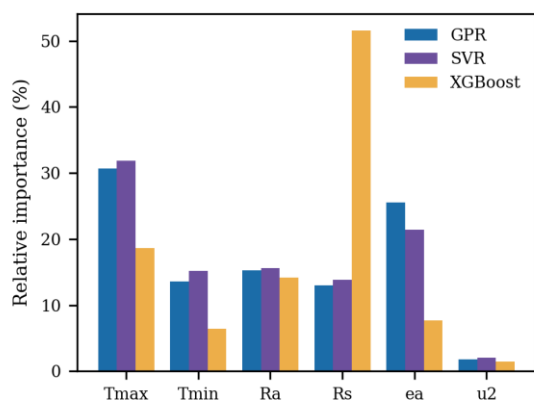


Figure 9. Normalised permutation importance of the inputs for combination C6

4. Discussion

The results agree with earlier findings that uncalibrated empirical equations can be strongly biased outside the climate for which they were developed (Tabari et al., 2013). Hargreaves–Samani, derived for semi-arid conditions, overestimated ET₀ in this humid climate by about 15%. The Makkink coefficient of 0.65, which approximates the Dutch operational reference evaporation (de Bruin and Lablans, 1998), underestimated FAO-56 PM by about 15%. The two references thus differ systematically, and conversion between them requires a factor of about 1.07 plus an offset. After local calibration, the radiation-based Makkink equation was almost as accurate as the ML models that used the same information. This supports its continued use where radiation is measured.

Without radiation data, the data-driven models were clearly better than any calibrated temperature-based equation, and GPR was the best model. This is consistent with studies in which kernel-based methods performed well for ET₀ estimation with limited data (Kisi, 2016; Fan et al., 2018). GPR has two advantages for small monthly samples: its hyperparameters are learned from the marginal likelihood rather than by grid search, and its smooth prior matches the nearly deterministic target. The relatively poor performance of XGBoost differs from daily-scale studies with thousands of samples (Fan et al., 2018). It reflects the small training sets and the inability of trees to represent smooth functions without many splits.

The high value of vapour pressure information is specific to the humid climate. With a mean relative humidity of 77%, the vapour pressure deficit is small and relatively variable, and the aerodynamic term of the PM equation responds strongly to it, as the sensitivity analysis showed. At stations that record temperature and humidity but not radiation, which is common, ML models can therefore approach full-PM accuracy. The significant increases in temperature, vapour pressure and wind speed at Wageningen during 1976–1999 largely offset one another in the ET₀ trend. This agrees with the view that ET₀ trends cannot be inferred from warming

alone (Roderick and Farquhar, 2002; McVicar et al., 2012).

The study has limitations. It used one station and a 24-year record, and the target was a computed PM value rather than measured evapotranspiration, so the models learn an equation rather than a physical flux. The early-morning vapour pressure recorded at the station was used as the daily mean *ea*, which may bias PM slightly. The transferability of the calibrated equations and ML models to other stations was not tested. Future work should include multiple stations and spatial cross-validation.

5. Conclusions

Monthly FAO-56 PM-ET₀ at Wageningen (1976–1999) was estimated with six empirical equations and five data-driven models under six input combinations using blocked cross-validation. The main conclusions are:

(1) Annual ET₀ averaged 631 mm and increased by 4.0 mm yr⁻¹ ($p = 0.051$), while temperature, vapour pressure and wind speed increased significantly. ET₀ was most sensitive to vapour pressure and maximum temperature.

(2) Uncalibrated empirical equations had RMSEs of 9–14 mm month⁻¹ and biases of up to $\pm 15\%$. Local calibration was essential, and the calibrated Makkink equation was the best empirical method (RMSE 4.35 mm month⁻¹).

(3) GPR was the best data-driven model for all input sets. With temperature only it reached RMSE 5.3 mm month⁻¹, better than calibrated Hargreaves–Samani (6.1). With temperature and vapour pressure it reached 3.1, and with all inputs 1.4 mm month⁻¹.

(4) In this humid climate, vapour pressure was the most informative single addition to temperature for the non-linear models. XGBoost was the weakest non-linear model for the small monthly samples.

Data Availability Statement

The Wageningen (Haarweg) daily weather files are distributed with the open-source PCSE package (<https://pypi.org/project/pcse/>). The processed monthly dataset and scripts are available from the corresponding author on request.

Conflict of Interest

The authors declare no conflict of interest.

References

- Allen, R.G., Pereira, L.S., Raes, D., & Smith, M. (1998). Crop evapotranspiration: Guidelines for computing crop water requirements. FAO Irrigation and Drainage Paper 56. FAO, Rome.
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794. <https://doi.org/10.1145/2939672.2939785>
- de Bruin, H.A.R., & Lablans, W.N. (1998). Reference crop evapotranspiration determined with a modified

- Makkink equation. *Hydrological Processes*, 12(7), 1053–1062.
- Fan, J., Yue, W., Wu, L., Zhang, F., Cai, H., Wang, X., Lu, X., & Xiang, Y. (2018). Evaluation of SVM, ELM and four tree-based ensemble models for predicting daily reference evapotranspiration using limited meteorological data in different climates of China. *Agricultural and Forest Meteorology*, 263, 225–241. <https://doi.org/10.1016/j.agrformet.2018.08.019>
- Feng, Y., Cui, N., Gong, D., Zhang, Q., & Zhao, L. (2017). Evaluation of random forests and generalized regression neural networks for daily reference evapotranspiration modelling. *Agricultural Water Management*, 193, 163–173. <https://doi.org/10.1016/j.agwat.2017.08.003>
- Ferreira, L.B., da Cunha, F.F., de Oliveira, R.A., & Fernandes Filho, E.I. (2019). Estimation of reference evapotranspiration in Brazil with limited meteorological data using ANN and SVM – A new approach. *Journal of Hydrology*, 572, 556–570. <https://doi.org/10.1016/j.jhydrol.2019.03.028>
- Gong, L., Xu, C.-Y., Chen, D., Halldin, S., & Chen, Y.D. (2006). Sensitivity of the Penman–Monteith reference evapotranspiration to key climatic variables in the Changjiang (Yangtze River) basin. *Journal of Hydrology*, 329, 620–629. <https://doi.org/10.1016/j.jhydrol.2006.03.027>
- Granata, F. (2019). Evapotranspiration evaluation models based on machine learning algorithms—A comparative study. *Agricultural Water Management*, 217, 303–315. <https://doi.org/10.1016/j.agwat.2019.03.015>
- Hamed, K.H., & Rao, A.R. (1998). A modified Mann–Kendall trend test for autocorrelated data. *Journal of Hydrology*, 204, 182–196.
- Hargreaves, G.H., & Samani, Z.A. (1985). Reference crop evapotranspiration from temperature. *Applied Engineering in Agriculture*, 1(2), 96–99.
- Kisi, O. (2016). Modeling reference evapotranspiration using three different heuristic regression approaches. *Agricultural Water Management*, 169, 162–172. <https://doi.org/10.1016/j.agwat.2016.02.026>
- Kumar, M., Raghuwanshi, N.S., Singh, R., Wallender, W.W., & Pruitt, W.O. (2002). Estimating evapotranspiration using artificial neural network. *Journal of Irrigation and Drainage Engineering*, 128(4), 224–233.
- Makkink, G.F. (1957). Testing the Penman formula by means of lysimeters. *Journal of the Institution of Water Engineers*, 11, 277–288.
- McCuen, R.H. (1974). A sensitivity and error analysis of procedures used for estimating evaporation. *Water Resources Bulletin*, 10(3), 486–497.
- McVicar, T.R., Roderick, M.L., Donohue, R.J., Li, L.T., Van Niel, T.G., Thomas, A., et al. (2012). Global review and synthesis of trends in observed terrestrial near-surface wind speeds: implications for evaporation. *Journal of Hydrology*, 416–417, 182–205. <https://doi.org/10.1016/j.jhydrol.2011.10.024>
- Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D., & Veith, T.L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE*, 50(3), 885–900.
- Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F., & Loumagne, C. (2005). Which potential evapotranspiration input for a lumped rainfall–runoff model? Part 2—Towards a simple and efficient potential evapotranspiration model for rainfall–runoff modelling. *Journal of Hydrology*, 303, 290–306.
- Priestley, C.H.B., & Taylor, R.J. (1972). On the assessment of surface heat flux and evaporation using large-scale parameters. *Monthly Weather Review*, 100(2), 81–92.
- Rasmussen, C.E., & Williams, C.K.I. (2006). *Gaussian Processes for Machine Learning*. MIT Press, Cambridge, MA.
- Roderick, M.L., & Farquhar, G.D. (2002). The cause of decreased pan evaporation over the past 50 years. *Science*, 298, 1410–1411. <https://doi.org/10.1126/science.1075390>
- Tabari, H., Grismer, M.E., & Trajkovic, S. (2013). Comparative analysis of 31 reference evapotranspiration methods under humid conditions. *Irrigation Science*, 31, 107–117. <https://doi.org/10.1007/s00271-011-0295-z>
- Thornthwaite, C.W. (1948). An approach toward a rational classification of climate. *Geographical Review*, 38(1), 55–94.
- Turc, L. (1961). Évaluation des besoins en eau d'irrigation, évapotranspiration potentielle. *Annales Agronomiques*, 12, 13–49.
- Vapnik, V.N. (1995). *The Nature of Statistical Learning Theory*. Springer, New York.