hydrology

MDPI

Article

Provenance-Aware Evaluation of Reference Evapotranspiration Methods at a Humid Tropical Site Using Random Forest and ERA5-Seamless Reconstruction

Christian José Mendoza Castiblanco [1,*], Andrés Javier Peña Quiñones [2] and Andrés Felipe Rodríguez Vásquez [1]

[1] Civil and Agricultural Department, Faculty of Engineering, Universidad Nacional de Colombia, Bogotá 111321, Colombia; afrodriguezv@unal.edu.co
[2] Corporación Colombiana de Investigación Agropecuaria-AGROSAVIA C.I.T la Libertad, Villavicencio 500008, Colombia; apena@agrosavia.co
* Correspondence: cjmendozac@unal.edu.co

Abstract

Assessing reference evapotranspiration (ETo) in humid tropical areas is often difficult because meteorological records are incomplete. This study compared four simplified methods with the FAO-56 Penman–Monteith (PM) method, depending on the source of meteorological data at the La Libertad station in the Colombian Orinoquía between 2018 and 2025. We checked hourly records for quality and filled missing data using a multistage approach that combined Random Forest and ERA5-Seamless reanalysis. Of 2713 complete days, only 527 (19.4%) were fully quality-checked, and none were from 2021 to 2024. In blocked validation, the hybrid reconstruction lowered hourly errors by 5–15% compared to calendar-based estimates and by 10–42% compared to raw reanalysis. The average PM ETo was 3.78 mm/day. The Priestley–Taylor method, when using the same net radiation as PM, matched daily PM results most closely (RMSE = 0.54 mm/day; 95% CI 0.49–0.60; PBIAS = +6.5%). Hargreaves–Samani and a daily Thornthwaite adaptation gave values about 20% higher than PM, and the simplified Blaney–Criddle was 44% higher. Monthly, Priestley–Taylor and conventional Thornthwaite gave nearly the same results (RMSE = 14.8 and 14.9 mm/month). The daily ranking of methods was the same on fully quality-checked days and when ETo was calculated from either observed or reconstructed data, though reconstructed data led to daily PM errors of about 1 mm/day. Priestley–Taylor should be considered only as a preliminary alternative when radiation, temperature, and humidity data are available, and further local validation is needed.

Keywords: FAO-56 Penman–Monteith; Priestley–Taylor; Hargreaves–Samani; meteorological gap filling; random forest; ERA5-seamless; data provenance; Colombian Orinoquía

Check for updates

Academic Editors: Adilson Pacheco De Souza, Frederico Terra De Almeida and Daniela Roberta Borella

Received: 23 October 2026
Revised: 16 November 2026
Accepted: 17 November 2026
Published: 20 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.

1. Introduction

Accurate estimation of reference evapotranspiration (ETo) is critical for effective agricultural water management, irrigation planning, and sustainable water use. Inadequate or excessive irrigation harms crop productivity, soil health, and nutrient availability, ultimately reducing irrigation efficiency [1,2]. ETo measures how much water the atmosphere would evaporate from a well-watered reference surface. It is a key factor in estimating how much water crops need and in planning irrigation systems [3,4]. Therefore, accurate ETo estimates are particularly important in agricultural regions, where fluctuations in meteorological conditions directly influence irrigation requirements and water resource availability.

Researchers have developed several empirical and physically based methods to estimate ETo for different climates and data availability. Among these, the Food and Agriculture Organization's FAO-56 Penman–Monteith equation is widely accepted as the standard because it includes the main physical processes that control water vapor transfer between the reference surface and the atmosphere [4,5]. However, applying it requires meteorological data on solar radiation, air temperature, atmospheric humidity, and wind speed, which agricultural monitoring networks do not always provide [4,6]. In these situations, methods that require fewer inputs, such as Hargreaves-Samani, Priestley-Taylor, Blaney-Criddle, and Thornthwaite, offer other ways to estimate atmospheric evaporative demand. However, how well they work depends on local climate, the calculation time scale, and the assumptions each method uses [4–7].

Researchers have investigated the suitability of simplified evapotranspiration methods across different agricultural and climatic environments, such as [8], which compared the Hargreaves–Samani method with FAO Penman–Monteith in northwestern São Paulo, Brazil, highlighting the potential of simplified approaches when complete meteorological information is unavailable. Similarly, ref. [9] compared several empirical methods in São Luís, Maranhão, emphasizing the need to evaluate their performance in specific regional climates. Studies in Colombia have further shown the importance of assessing evapotranspiration estimation methods for local agricultural applications [3,10]. These findings show that choosing the right method should rely on regional performance assessments, not just on how widely an empirical equation can be applied. Also, when comparing evapotranspiration estimates, consider the original time scales of some temperature-based methods. For example, Thornthwaite was designed to estimate monthly potential evapotranspiration [7], whereas subsequent adaptations have extended its application to daily reference evapotranspiration estimation [11].

Beyond choosing an appropriate estimation method, the availability and quality of meteorological observations pose significant challenges for reliable ETo assessment. Malfunctions, sensor deterioration, maintenance gaps, and incomplete records often disrupt meteorological time series, limiting methods that rely on multiple atmospheric variables. Previous studies have addressed missing climatic data using statistical and computational approaches, such as artificial neural networks, satellite-derived information, and optimization algorithms [12,13]. However, long gaps in observations and missing variables make things more difficult, especially when the available predictors do not capture short-term weather changes well.

Recent advances in machine learning (ML) have enhanced evapotranspiration estimation using limited meteorological data. Techniques such as linear regression, artificial neural networks, support vector machines, random forests, and deep learning predict ETo using fewer climate variables. The accuracy of these methods depends on climate, predictor selection, and model configuration [10,14–20]. Predicting ETo directly with machine learning is different from first reconstructing meteorological variables for use in physical or empirical equations. Traditional approaches need to keep the changes and consistency in temperature, humidity, solar radiation, and wind speed to produce reliable ETo estimates. Products like ERA5 offer consistent atmospheric data to support station records, but it is important to check their local accuracy and uncertainties [21].

Among ML algorithms, Random Forest (RF) [22] is particularly effective for reconstructing meteorological inputs at a single station. As an ensemble of regression trees, it captures nonlinear responses and interactions, such as the dependence of relative humidity on temperature, cloudiness, and time of day, without requiring a predefined functional form. It accommodates various predictor types, including continuous station and reanalysis variables, cyclic calendar encodings, and binary missingness indicators. Because tree splits

are unaffected by monotonic transformations, predictors do not need to be scaled. Random Forest is also robust to correlated and partially redundant predictors and requires tuning only a few key hyperparameters, which helps prevent overfitting when training data are temporally clustered. Finally, RF-based imputation has performed well for mixed-type data with complex missingness patterns [23], similar to station records where sensors may fail individually or together. However, these strengths do not guarantee superiority, so we compared RF with k-nearest neighbors, regularized regression, climatological medians, raw reanalysis, and a linear bias correction of the reanalysis.

These factors matter especially in the Colombian Orinoquía. The area's humid tropical climate, changes in cloud cover and sunlight, and seasonal shifts in the atmosphere can all influence how evapotranspiration is measured. The region also faces broader hydroclimatic changes linked to the El Niño–Southern Oscillation (ENSO), but their impact can vary by location and time of year [24]. Consequently, evaluating ETo over an extended period requires not only consistent meteorological information but also examining temporal variability and the estimates' sensitivity to the underlying meteorological inputs.

Previous studies in Colombia have compared ETo equations using station records from the Cauca River Valley [3] or predicted ETo with machine-learning models using limited predictors [10,15]. However, less attention has focused on how the choice of ETo equation depends on the source of meteorological data, especially when station records have many gaps. This study offers a provenance-aware evaluation of ETo methods. We measure how well four simplified equations—Hargreaves–Samani, Priestley–Taylor, Blaney–Criddle, and Thornthwaite—match the FAO-56 Penman–Monteith method in the humid tropics of the Colombian Orinoquía. We also test if the ranking of these methods stays consistent when using different levels of observational support, such as fully observed, partially reconstructed, and reanalysis-assisted days. The hybrid Random Forest–ERA5 reconstruction is an important step and is compared to simpler alternatives but is not the main focus of our analysis. Using meteorological data from the La Libertad station in Villavicencio from 2018 to 2025, our specific objectives are to (i) reconstruct hourly meteorological inputs with clear provenance labels and independent blocked validation; (ii) evaluate how well the four methods agree with PM at daily and monthly scales, including uncertainty intervals that consider temporal dependence; and (iii) check if the method ranking is stable across different data-provenance scenarios. We also briefly describe ETo across different El Niño–Southern Oscillation (ENSO) phases for context. Since our analysis uses only one station and takes PM as the reference instead of measured evapotranspiration, our results are meant to guide method selection at similar sites and highlight when independent validation is needed.

2. Materials and Methods

2.1. Study Area and Climate Characterization

We obtained meteorological data to estimate reference evapotranspiration from the La Libertad automated weather station, installed by the Institute of Hydrology, Meteorology and Environmental Studies (IDEAM) at the La Libertad Experiment Station. This station is part of the Colombian Corporation for Agricultural Research (AGROSAVIA) network. It is located in Villavicencio, Department of Meta, in the Colombian Orinoquía (4.0574° N, 73.4679° W; about 340 m above sea level), at the transition between the Eastern Cordillera foothills and the Orinoquía plains (Figure 1a). We downloaded records through the Hydrological and Meteorological Information System (DHIME), administered by IDEAM. The study area has humid tropical conditions because of its proximity to the Andean foothills and the Eastern Plains. It is mainly classified as a tropical rainforest climate (Af) with some transitional monsoonal features (Am). Air temperatures remain high year-round,

while precipitation, cloud cover, and solar radiation vary seasonally. The wet season typically lasts from April to November, and drier conditions prevail from December to March. ERA5-Seamless precipitation data for 2019–2025 indicate a unimodal pattern, with peak rainfall in May and a minimum from December to February (Figure 1b). The mean annual ERA5-Seamless total (approximately 1510 mm) is, however, only about one-third of the 1981–2010 gauge-based normal at Villavicencio–Vanguardia airport, approximately 20 km northwest of the station (≈4540 mm) [25], although both show the same seasonal regime, with maximum rainfall in May and minimum rainfall in January. This underestimation is consistent with the tendency of coarse reanalyses to underestimate convective rainfall in foothill regions. Consequently, Figure 1b is referenced solely to illustrate the seasonal precipitation regime. These climatic conditions, combined with hydroclimatic variability linked to the El Niño–Southern Oscillation (ENSO), create a pertinent context for evaluating evapotranspiration methods [24]. In this setting, persistent atmospheric humidity and seasonal fluctuations in available energy may influence the relative importance of radiative and aerodynamic processes in determining reference evapotranspiration (ETo). Therefore, without assessing method agreement under local conditions, the climatic applicability of approaches developed for other regions cannot be assumed [3–6].

2.2. Meteorological Data Acquisition and Quality Control

The dataset comprised hourly weather observations from 28 July 2018 to 31 December 2025. The original file had 50,805 records with missing values and breaks in the data. We created a continuous hourly index for the study period, which yielded 65,120 expected timestamps. The variables were minimum, mean and maximum air temperature (Tmin, Tmean and Tmax; °C), relative humidity (RH; %), global solar irradiance (Rs; W/m^2) and wind speed (u; m/s). A visible-radiation variable was also examined during the quality checks, but it was not used in the ETo calculations because its measurements were inconsistent. Quality control (QC) involved screening for missing timestamps, invalid readings, physically inadmissible values, and anomalous temporal patterns. We examined temperature measurements for implausible values and violations of $Tmin \leq Tmean \leq Tmax$. We restricted relative humidity values to 0–100% and flagged negative or unusually high wind speeds. We examined radiation records for nighttime inconsistencies, persistent daytime zeros, and changes indicating instrument malfunction. For Rs, negative values were rejected, whereas zero values were retained when they occurred at night or at sunrise (ERA5-Seamless shortwave radiation ≤ 3 W/m^2), because zero irradiance is physically expected at those hours; persistent daytime zeros were treated as evidence of instrument malfunction. The quality assessment identified an extended quarantine period for global solar radiation from 19 October 2021 to 16 November 2024, and anomalous behavior of the visible-radiation variable extending to 20 August 2025. Therefore, we excluded suspect observations from model training rather than treating them as valid measurements. The evidence supporting the quarantine is described below.

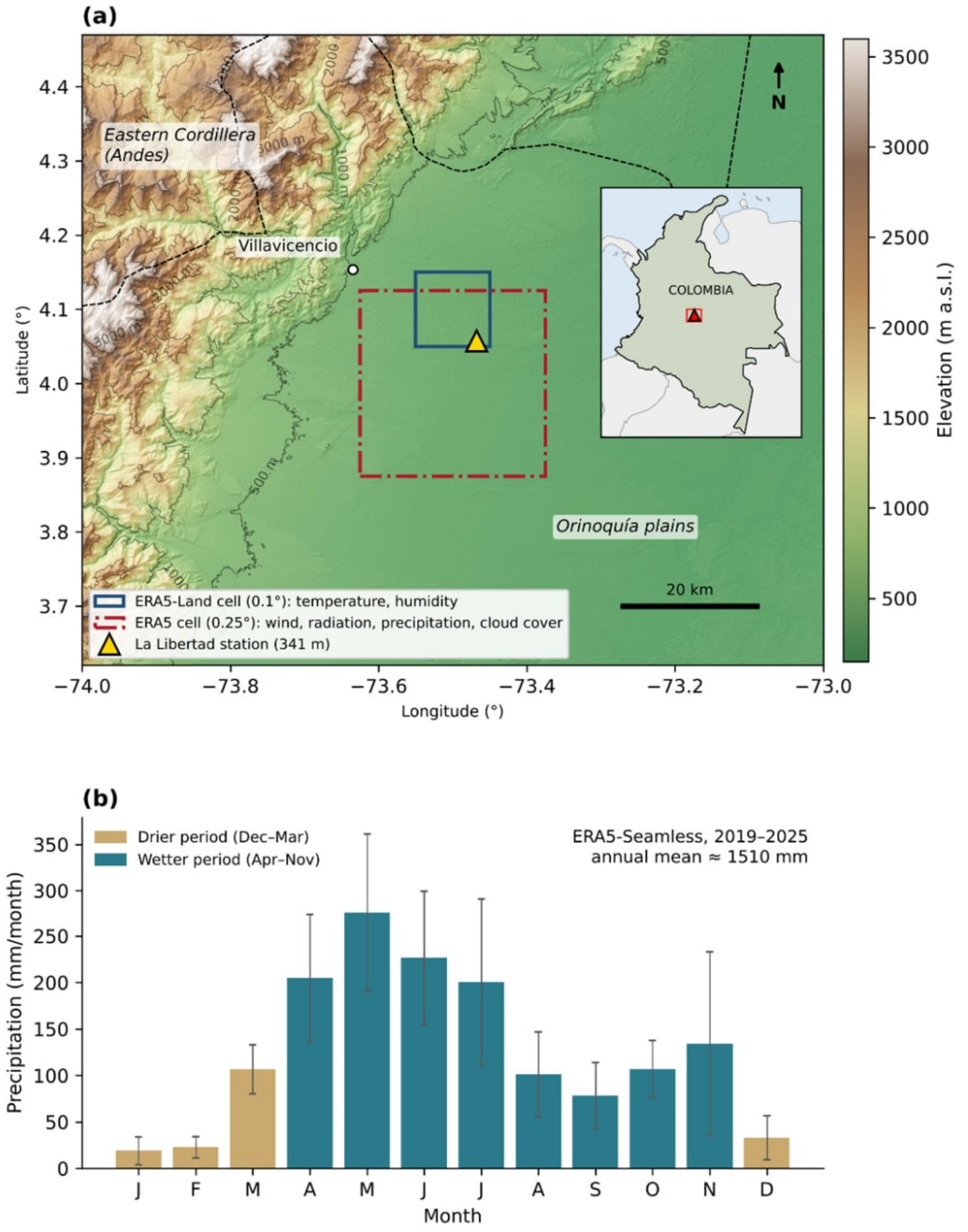


Figure 1. (**a**) Location of the La Libertad automated weather station (4.0574° N, 73.4679° W; 341 m a.s.l. in the DEM), Villavicencio, Meta, on a 3-arc-second digital elevation model (DEM) built from ASTER GDEM version 3 north of 4° N [26] and SRTM south of 4° N [27], with hillshading, 500, 1000, 2000 and 3000 m contours and departmental boundaries (dashed lines; Natural Earth). The Eastern Cordillera foothills lie to the west and the Orinoquía plains to the east. Boxes show the ERA5-Land (0.1°) cell used for temperature and humidity and the ERA5 (0.25°) cell used for wind, solar radiation, precipitation and cloud cover. Inset: location in Colombia. (**b**) Mean monthly ERA5-Seamless precipitation at the station grid cells, 2019–2025; error bars show the interannual standard deviation.

We based the radiation quarantine on the dataset's overall behavior, not on a fixed threshold. Between 19 October 2021 and 16 November 2024, 13,889 hourly global-radiation values were recorded. Of these, 12,250 (88.2%) were exactly zero, including 6197 of the 7225 daytime records (85.8%; daytime defined by ERA5-Seamless shortwave radiation > 0 W/m^2). For comparison, ERA5-Seamless radiation during the same period averaged 215 W/m^2, similar to the pre-quarantine (208 W/m^2) and post-quarantine

(201 W/m^2) values (Table S2; Figure S1). These zero values therefore indicate a sensor or data-logger failure rather than actual atmospheric conditions; no maintenance record documenting the cause was available. We excluded the quarantined values rather than correcting them because a constant zero output carries no recoverable signal, no co-located reference pyranometer or calibration record was available to derive a transfer function, and any correction would have required the same station–reanalysis relationship that the reconstruction later exploits. Outside the quarantine, all 18,906 available pre-quarantine hours (July 2018–October 2021) passed QC. Of the 8885 post-quarantine hours (November 2024–December 2025), the 641 negative values (−1 to −2 W/m^2, all at night) were rejected, and 8244 values (92.8%) were accepted. The 809 zero values of this period were retained: 806 occurred at night, and the remaining three occurred at 06:00 local time, when ERA5-Seamless radiation was only 1–3 W/m^2. Thus, the radiation models were fitted with 27,150 QC-accepted hourly Rs values (18,906 before and 8244 after the quarantine; Table S1). Nighttime values of the accepted records were small positive offsets before the quarantine (mean 4.8 W/m^2) and zeros or small offsets after it (mean 1.8 W/m^2). They were retained without correction because they contribute only 0.20 and 0.09 MJ/m^2/day, respectively (≈1% of daily Rs), which corresponds to ≈0.04–0.06 mm/day in the radiative terms of PM and PT. Comparability between the two periods was assessed with these QC-accepted values: the regression slope between station and ERA5-Seamless radiation was 0.83 before and 0.81 after the quarantine using all hours, and 0.78 and 0.76 using daytime hours only, a difference of about 3%, indicating that the two periods are comparable in magnitude (Table S2).

We kept the original station records, the QC-accepted measurements and the reconstruction status flags all separate in order to maintain data provenance. We also classified the missing values by length of absence and by the availability of valid auxiliary meteorological measurements. A complete daily record was one in which there were valid values for all six variables in each of the 24 h; we excluded the first incomplete day from the daily ETo calculation. The resulting QC database was not treated as a temporally representative subsample of the full period because data availability varied greatly from year to year. Instead, the QC observations supplied the model-training targets and an observation-only comparison scenario, and the meteorological reconstructions were given distinct origin flags for further analysis.

2.3. Meteorological Data Gap Filling

We implemented a multi-stage reconstruction procedure to deal with isolated sensor failures and with periods of simultaneous data loss. The approach involved a combination of multivariate Random Forest (RF) models, calendar-based predictors and ERA5 Seamless reanalysis (Table 1). RF was chosen as it can capture non-linear relationships and interactions between the various meteorological variables and can handle a mixture of predictor types and missingness indicators (Section 1). This is in line with the use of machine learning in the context of incomplete climatic data and evapotranspiration estimation [10,12,13,19,20]. The individual stages of the reconstruction were assessed separately as the performance of a model with observed auxiliary variables cannot be assumed to be the same in the case of a complete station outage. Likewise, the use of a continuous external atmospheric field does not necessarily remove any potential local bias and it will still need to be statistically linked to the accepted station observations. The staged approach was designed to maintain the integrity of the measurements, to make use of local cross-variable information where possible and to use external predictors where the station data alone would not be sufficient to fill the gaps.

Table 1. Input requirements of each ETo implementation as applied in this study.

Method	Timescale	Tmax, Tmin	Tmean	RH	Rs	u_2	Other Inputs
FAO-56 PM [4]	Daily	•	— [a]	•	•	•	z, φ, J (Ra, Rso, γ)
Hargreaves–Samani [4,6]	Daily	•	— [a]	—	—	—	φ, J (Ra)
Priestley–Taylor	Daily	•	— [a]	• [c]	•	—	z, φ, J (Ra, Rso, γ); α = 1.26
Blaney–Criddle, simplified [4,6]	Monthly; daily extension	—	•	—	—	—	φ, J (p)
Thornthwaite [7]	Monthly	—	• (monthly)	—	—	—	φ, J (N); I, a
Thornthwaite–Pereira–Pruitt [11]	Daily	•	• [b]	—	—	—	φ, J (N); I, a; k = 0.69

• required; — not required. [a] Computed as (Tmax + Tmin)/2. [b] Tavg in Equation (15). [c] Through actual vapor pressure in the FAO-56 net longwave term; PT therefore requires humidity in this implementation. φ, latitude; J, day of year; z, elevation; Ra, extraterrestrial radiation; Rso, clear-sky radiation; N, daylight duration; p, daily percentage of annual daylight hours.

2.4. Preliminary Model Assessment and Station-Based RF Reconstruction

Candidate gap-filling methods included RF, k-nearest neighbors, regularized regression, and month–hour climatological medians. We assessed performance by withholding valid QC observations and reconstructing them under conditions representing either isolated missing variables or simultaneous station outages. Temporal predictors captured the diurnal and annual cycles. We included observed auxiliary meteorological variables only when QC accepted them. Calendar-based estimates provided reference predictions but tended to suppress short-term variability during extended gaps. We evaluated candidate methods by variable rather than choosing a universal algorithm because the magnitude and temporal structure of temperature, humidity, radiation, and wind differed. We designed separate tests for missing radiation values with the remaining predictors available and for uninterrupted multivariable gaps. The latter were especially important in evaluating how model skill changed as observation-based predictors were removed. For station-based multivariate reconstruction, we fitted a separate RF model for each target variable using available QC auxiliary measurements, harmonic calendar features, and missingness indicators where applicable. We applied median imputation to missing predictor entries when prediction required at least two valid auxiliary variables, including at least one temperature measurement. The RF configuration comprised 60 trees ($n_{estimators} = 60$), $\left(max_{depth} = 14\right)$, $\left(min_{samples\ leaf} = 3\right)$, and $(random_{state} = 42)$. We selected the number of trees through a sensitivity analysis rather than exhaustive optimization: we evaluated the out-of-bag (OOB) error for 20, 40, 60, 70, 100, 200, and 300 trees for each target variable (Table S1). The OOB RMSE decreased by 2–5% between 20 and 60 trees and changed little thereafter; increasing the ensemble from 60 to 300 trees reduced it by only 0.8–1.4%, depending on the variable. This behavior is consistent with the general finding that RF performance converges with a moderate number of trees, beyond which additional trees mainly increase computational cost [28]; we fixed the maximum depth and minimum leaf size a priori to limit overfitting to the temporally clustered QC observations.

We assigned provenance flags to the estimated values according to gap length: estimates for gaps of 168 h or less were classified as short-gap reconstructions, and those for longer gaps as exploratory long-gap estimates. Short gaps (up to 24 h) with insufficient auxiliary observations could be filled with temporal alternatives (climatological medians, regularized regression for RH and calendar-only RF for wind speed). The data-provenance scenarios were defined cumulatively at the daily scale. The conservative scenario (950 days) comprised the fully QC-observed days (527) and the days completed with eligible short-gap RF estimates (423). The exploratory scenario (1511 days) additionally comprised the days completed with long-gap RF estimates (528) and 33 days (6 in 2019 and 27 in 2025) whose

short simultaneous gaps (typically 1–2 h and at most 7 h per day) had been filled with the temporal alternatives. In the final complete series, these short temporal estimates were replaced by hybrid RF–ERA5 predictions (Section 2.6), so the 33 days are counted as RF–ERA5 days in Table 2; this is why the exploratory total exceeds 527 + 423 + 528 = 1478 days. All exploratory-scenario statistics were computed on these 1511 days. The intermediate versions were kept separate from the final complete series so that exploratory estimates could not be interpreted as observations, and the long-gap RF stage was kept as a distinct provenance class for estimates that passed the relevant plausibility checks.

2.5. ERA5-Seamless Reanalysis Data

For intervals with insufficient station measurements, hourly predictors were retrieved through the Open-Meteo Historical Weather API [29] using the ERA5-Seamless product. This product merges temperature and humidity from ERA5-Land (0.1°, ≈9 km) [30] with wind, solar radiation, precipitation and cloud cover from ERA5 (0.25°, ≈28 km) [21]. Accordingly, temperature and humidity correspond to the ERA5-Land cell centered at approximately 4.10° N, 73.50° W (337 m model elevation), about 5.9 km from the station (≈4.7 km north and ≈3.6 km west), with an elevation difference of only ≈3 m; consequently, no lapse-rate adjustment was applied. Wind, radiation, precipitation and cloud cover correspond to the nearest ERA5 grid point (4.00° N, 73.50° W; ≈7 km from the station), whose cell represents an area of roughly 28 km × 28 km (Figure 1a). Six hourly variables were downloaded: 2-m air temperature (temperature_2m), 2-m relative humidity (relative_humidity_2m), surface shortwave radiation (shortwave_radiation), 10-m wind speed (wind_speed_10m), precipitation, and total cloud cover (cloud_cover). Records were aligned with the 65,120 station timestamps in the America/Bogota time zone (UTC−5), using temperature in °C, wind speed in m/s, and irradiance in W/m^2.

We compared the downloaded reanalysis values with coincident QC station measurements to assess consistency and local representativeness. Therefore, we used the reanalysis variables as predictors rather than direct station observations. In particular, the 10-m reanalysis wind variable was distinguished from the station wind measurements used at 2 m for the FAO-56 calculation. Precipitation and cloud cover were included as supplementary predictors of local atmospheric conditions but were not required directly by the selected daily FAO-56 equation.

Topography limits the local representativeness of the grid-cell values. The station lies on the Orinoquía plains, east of the Eastern Cordillera foothills. According to the DEM (Figure 1a), terrain elevation rises from 341 m at the station to 500 m about 21 km to the northwest, 1000 m at about 24 km, 2000 m at about 31 km and 3000 m at about 38 km. Both reanalysis cells containing the station lie entirely on the plains (DEM elevations of 296–373 m in the ERA5-Land cell and 255–443 m in the ERA5 cell), whereas the reanalysis orography smooths the steep transition to the Andes. At this transition, orographic lifting, foothill–plain circulations and nocturnal convection generate precipitation and cloud gradients at scales finer than the 9–28 km grid. These effects mainly affect precipitation and cloud cover and, therefore, incoming shortwave radiation—which, like wind, is taken from the coarser 0.25° ERA5 grid—and, to a lesser extent, near-surface humidity and wind, which are also influenced by local surface roughness and exposure. For this reason, we used the reanalysis fields only as predictors, learned from their relationship with the station using coincident QC observations, and quantified their raw agreement with the station separately for each variable (Section 3.1; Table 3).

2.6. Hybrid RF–ERA5 Reconstruction and Physical Constraints

We trained separate hybrid RF models for Tmin, Tmean, Tmax, RH, Rs, and wind speed using QC station measurements as targets. The models shared 23 predictors: four pairs of sine–cosine harmonics describing the hour and day of year (eight predictors), the six ERA5-Seamless variables, calendar year, and eight lagged/leading reanalysis values (one hour before and after for temperature, relative humidity, radiation, and cloud cover). The fitted RF parameters were n_estimators = 70, max_depth = 15, min_samples_leaf = 4, max_features = 0.8, random_state = 42, and n_jobs = −1, selected with the same sensitivity procedure described in Section 2.4 (Table S1). We fitted models independently for each target. When predicting missing or low-confidence calendar-based values, the hybrid models used only external (ERA5-Seamless) and calendar predictors; no station observations entered the predictor set. Consequently, during the longest outages—including the radiation quarantine (19 October 2021–16 November 2024) and the periods in which all station sensors were simultaneously unavailable—the reconstruction was purely ERA5-Seamless- and calendar-based, and station information influenced these estimates only through the relationships learned from QC observations recorded in other periods. Removing the calendar-year predictor did not improve performance when entire years were withheld (Section 3.1; Table S4), so it was retained. The final hourly dataset consisted of the QC-accepted observations and the selected estimates from the station-based RF stages. The remaining gaps, as well as the calendar-only and other short temporal estimates, were filled with hybrid RF–ERA5 predictions. The reconstructed values were constrained to physically plausible ranges: 10–45 °C for temperature, 0–100% for RH, 0–30 m/s for wind speed and 0–1400 W/m^2 for solar irradiance. Nighttime reconstructed radiation was set to zero if both the local hour and the reanalysis radiation were indicative of no sunlight. To maintain thermal ordering (Tmin ≤ Tmean ≤ Tmax), we only adjusted the reconstructed values and did not alter the accepted QC observations. A boundary residual correction was tried but was not included in the final reconstruction as it did not consistently improve the validation.

The final series retains variable-specific flags that distinguish QC observations, short-gap RF, exploratory long-gap RF and RF–ERA5 estimates. This is particularly important for days on which only a subset of the variables was reconstructed. A day is classified as fully QC only if all six variables were observed and accepted for the full 24 h period. The QC, conservative, exploratory and complete reconstruction scenarios are all carried forward to assess sensitivity to data provenance. A complete hourly series represents numerical coverage rather than uniform observational confidence.

2.7. Validation of Meteorological Reconstruction

Validation used QC observations excluded from model fitting. Initial withheld-observation tests assessed the candidate methods under isolated sensor-loss conditions. Longer contiguous blocks simulated simultaneous station outages and prevented the use of withheld station observations as predictors. The hybrid RF–ERA5 evaluation employed six blocked periods of 7–90 days (one of 7 days, four of 30 days and one of 90 days in 2018, 2019, 2022, 2024 and 2025), totaling 217 days or 5208 hourly timestamps (block lengths, years and block-level metrics in Table S3). We computed performance only where QC observations existed within these blocks. Thus, the number of evaluated observations differed among variables. We evaluated calendar-only reference models on the same withheld periods. For each validation fold, we refitted the complete pre-processing and modeling chain without the withheld block. We estimated the median values used to impute missing predictors, the month–hour climatologies, the regularized-regression coefficients, the station-based RF models, and the hybrid RF–ERA5 models exclusively from data outside the block. An

independent re-run of the complete chain within folds reproduced the original validation statistics to within 1% (Table 3). Therefore, no withheld QC value was available, directly or through an intermediate model, when the corresponding block was predicted. We used the independent blocked design in addition to the isolated-gap tests because random point-wise masking can provide overly favorable performance estimates when temporally adjacent measurements remain accessible. To assess whether the RF–ERA5 approach adds value over simpler alternatives, we compared it on the same withheld blocks against: (i) the raw ERA5-Seamless values with wind speed reduced from 10 m to 2 m using the FAO-56 logarithmic profile ($u_2 = 0.748u_{10}$) [4]; (ii) a linear bias correction of each ERA5-Seamless variable, with the bias corrected by ordinary least squares to the coincident QC observations, separately for each calendar month; and (iii) the calendar-only RF reference. Robustness to longer outages was examined in the two years with the most QC observations (2019 and 2020) using a leave-one-year-out experiment in which all QC data from the excluded year were not used in the model fitting and the model was then applied to the entire year; in addition, contiguous blocks of 120 and 180 days starting on 1 March 2019 were withheld. The leave-one-year-out experiment was also carried out without the calendar-year predictor to check whether this predictor in fact degrades the extrapolation to years without training targets. Given that ETo is a function of all the inputs, we also evaluated the reconstruction in terms of its ability to maintain the inter-variable relationships: the Pearson correlation matrices of daily Tmax, Tmin, RH, Rs and wind speed derived from the reconstructed values within the withheld blocks were compared to those of the coincident QC observations. We also calculated the five ETo methods for the withheld days twice, once with the QC-observed inputs and once with the reconstructed inputs, to assess the error that was propagated into ETo and to see whether the relative ranking of the methods with respect to PM was preserved on those dates. Accuracy and bias were measured using mean absolute error (MAE), root mean square error (RMSE), mean bias error (MBE), and Pearson's correlation:

$$MAE = \left(\frac{1}{n}\right)\sum|P_i - O_i| \quad (1)$$

$$RMSE = \sqrt{\left[\left(\frac{1}{n}\right)\sum(P_i - O_i)^2\right]} \quad (2)$$

$$MBE = \left(\frac{1}{n}\right)\sum(P_i - O_i) \quad (3)$$

where P_i and O_i denote predicted and QC-observed values, respectively, and n is the number of evaluated observations. Solar radiation was also evaluated after aggregation to daily energy totals; the distribution and temporal variability of reconstructed values were compared with those of observed periods to identify excessive smoothing and physically implausible behavior. Model performance was interpreted jointly with the duration of each withheld interval and with the amount of valid QC data actually available within it; therefore the additional assessment of boundary adjustments was performed against withheld QC values before the final choice to omit the correction.

The completed database yielded 2713 days with all six hourly meteorological variables available. The corresponding QC, conservative, and exploratory scenarios contained 527, 950 (527 + 423), and 1511 (527 + 423 + 528 + 33; Section 2.4) complete days, respectively, and their annual distribution is reported in Table 2. The complete scenario was used for the principal ETo comparisons and the other scenarios for sensitivity analyses. The 5208 withheld hours represented validation windows rather than an equal number of QC observations for every variable.

2.8. Daily Aggregation and Atmospheric Parameters

Hourly data were aggregated into daily variables after reconstruction. The daily Tmin and Tmax were obtained from the lowest and highest hourly values, respectively. Daily mean temperature, RH, and wind speed were calculated as arithmetic hourly means. Assuming hourly solar-radiation values represent mean irradiance over each hour, daily global solar radiation was obtained by integrating the 24 hourly values:

$$R_{s,d} = 0.0036 \sum R_{s,h} \tag{4}$$

where $R_{s,d}$ is expressed in MJ/m^2/day and $R_{s,h}$ in W/m^2. Only full 24-h dates were retained and the distinction between the daily average of hourly mean temperature and the FAO-56 temperature definition was maintained for sensitivity assessment. All scenarios used the same aggregation rules, so that comparisons between ETo methods were not affected by inconsistent definitions of daily meteorological inputs.

Astronomical and atmospheric parameters were calculated using FAO-56 procedures [4]; extraterrestrial radiation (R_a) and maximum daylight duration (N) were computed from latitude and day of year, atmospheric pressure $(P; \text{kPa})$ and the psychrometric constant $(\gamma; \text{kPa/°C})$ were obtained from elevation z (m):

$$P = 101.3 \left[\frac{293 - 0.0065z}{293} \right]^{5.26} \tag{5}$$

$$\gamma = 0.000665\, P \tag{6}$$

Clear-sky radiation was estimated as $R_{so} = \left(0.75 + 2 \times 10^{-5} z\right) R_a$. For the FAO-56 calculation, daily mean air temperature was defined as $T = (Tmax + Tmin)/2$. Saturation vapor pressure (es) was calculated from Tmax and Tmin; actual vapor pressure (ea) was approximated from mean daily RH as $ea = es(RH/100)$. Saturation vapor pressure was calculated as $es = [e°(Tmax) + e°(Tmin)]/2$, where e°(T) = 0.6108 exp[17.27T/(T + 237.3)]. The slope of the saturation vapor pressure curve was obtained from $\Delta = 4098 e°(T)/(T + 237.3)^2$, using the FAO-56 mean temperature. Vapor pressure deficit $(es - ea)$ and net radiation were then calculated consistently across scenarios. This approximation for *ea* from mean RH was retained because the daily input dataset did not contain separately aggregated maximum and minimum relative humidities.

Net radiation was obtained from net shortwave radiation using albedo 0.23 and FAO-56 net longwave radiation:

$$R_n = R_{ns} - R_{nl} \tag{7}$$

The ratio R_s/R_{so} was capped at 1 in the longwave calculation; the unmodified FAO-56 longwave expression was used in the principal ETo estimates; a sensitivity analysis considered clipping negative net outgoing longwave-radiation estimates to zero. Daily soil heat flux was assumed negligible $(G = 0)$ [4]. The calculated radiation quantities were checked for physical plausibility, including consistency between observed solar radiation and estimated clear-sky radiation. The main estimates used the direct FAO-56 expression even where its empirical longwave term became negative under low ratios of R_s/R_{so}, allowing the sensitivity analysis to quantify rather than conceal the effect of the optional lower bound.

2.9. Reference Evapotranspiration Estimation Methods

We evaluated five methods: FAO-56 Penman–Monteith (PM), Hargreaves–Samani (HS), Priestley–Taylor (PT), Blaney–Criddle (BC) and Thornthwaite (T). Two temporal comparisons were made. In the conventional approach, PM, HS and PT were calculated daily and summed to monthly totals, whereas BC and T were calculated with their monthly formulations. The daily approach calculated all five methods, using the daily versions of

BC and T. Monthly comparisons were restricted to complete calendar months so that all methods covered the same period. The conventional and daily versions of BC and T were stored separately. ETo denotes the PM reference and comparable estimates, whereas the original Thornthwaite equation is referred to as potential evapotranspiration (PET) because of its distinct conceptual basis. Table 1 details the inputs required by each implementation. PT was implemented with the same net radiation as PM; because the FAO-56 net longwave term depends on actual vapor pressure, PT requires humidity data in addition to Rs, Tmax and Tmin, and it differs from PM mainly by omitting the aerodynamic term.

2.9.1. FAO-56 Penman–Monteith Method

FAO-56 Penman–Monteith was used as the standard reference [4,5]; its daily equation combines radiative and aerodynamic terms:

$$ETo\ PM = \frac{0.408\,\Delta(R_n - G) + \gamma\left(\frac{900}{(T+273)}\right)u_2(e_s - e_a)}{\Delta + \gamma(1 + 0.34u_2)} \quad (8)$$

where ETo is expressed in mm/day, R_n and G in MJ/m^2/day, T in °C, u_2 in m/s, es and e_a in kPa, and Δ and γ in kPa/°C. Station wind speed was treated as measured at the adopted 2 m reference height, without additional height conversion. Alternative calculations examined the effect of replacing FAO-56 mean temperature by the mean of hourly temperatures and of applying the alternative longwave-radiation treatment.

2.9.2. Blaney–Criddle Method

The simplified Blaney–Criddle formulation relates evapotranspiration to temperature and daylight duration [4,6]. The version evaluated here is the uncorrected expression, i.e., without the adjustment factor that the FAO-24 formulation applies as a function of minimum relative humidity, sunshine ratio and daytime wind speed [31]; the results therefore refer to this simplified version and not to Blaney–Criddle formulations that include those climatic adjustments:

$$ET\ BC = p(0.46T + 8) \quad (9)$$

where T is mean air temperature (°C) and p is the average daily percentage of annual daylight hours. For the conventional monthly method, mean monthly temperature and the mean daily daylight percentage were used to calculate an estimated daily rate, subsequently multiplied by the number of days in the month. The daily extension used the temperature and daylight percentage for each individual date, with p derived from daily astronomical daylight hours relative to the total for its calendar year. Daily estimates were summed to monthly totals for comparison with the conventional formulation. The daily version was treated as a temporal extension of the simplified equation rather than as a locally calibrated model. Because the formula is linear in temperature, differences between the two monthly totals principally depend on the within-month covariation between temperature and daily daylight percentage.

2.9.3. Hargreaves–Samani Method

The temperature-based Hargreaves–Samani method was calculated following FAO-56 guidance [4,6]:

$$ETo\ HS = 0.0023(T + 17.8)\sqrt{(Tmax - Tmin)(0.408R_a)} \quad (10)$$

where R_a is extraterrestrial radiation $\left(\text{MJ/m}^2\text{/day}\right)$ and T, Tmax, and Tmin are daily mean, maximum, and minimum temperatures (°C). The thermal amplitude was determined using daily extrema, rather than the averages of hourly Tmin and Tmax, to represent the actual daily temperature range required by the equation.

2.9.4. Priestley–Taylor Method

Priestley–Taylor estimates evapotranspiration from available energy using an empirical coefficient to account for aerodynamic effects under limited advection:

$$ET\ PT = \propto \left[\frac{\Delta}{\Delta + \gamma}\right] 0.408(R_n - G) \quad (11)$$

The coefficient α was fixed at 1.26, R_n, Δ, and γ were calculated consistently with the PM formulation. Because the FAO-56 net longwave term in Rn depends on actual vapor pressure, the PT implementation tested here requires Rs, Tmax, Tmin and RH, and differs from PM by omitting the aerodynamic term (wind speed and vapor pressure deficit); it is therefore not a temperature-and-radiation-only alternative (Table 1). The resulting estimates, expressed in mm/day, were aggregated to monthly totals for the conventional comparison.

2.9.5. Thornthwaite Method and Daily Adaptation

The conventional Thornthwaite method [7] estimates monthly potential evapotranspiration (PET) from temperature and a thermal index:

$$PET\ T = 16(10Tm/I)^a \quad (12)$$

$$I = \sum (Tm/5)^{1.514} \quad (13)$$

$$a = 6.75 \times 10^{-7} I^3 - 7.71 \times 10^{-5} I^2 + 0.017912\ I + 0.49239 \quad (14)$$

where Tm denotes the positive monthly mean temperature (°C), I is the annual thermal index, and a is its empirical exponent. Equation (12) represents a standard 30-day month with a 12-h photoperiod and was adjusted by monthly day length and calendar-month duration. For consistency among meteorological scenarios, I was derived from monthly mean temperatures of the complete series over 2019–2025 (I = 144.5), constituting a seven-year reference rather than a 30-year climate normal. The adopted thermal function included a polynomial high-temperature correction for T ≥ 26.5 °C [32], expressed as $PET = -415.85 + 32.24T - 0.43T^2$ for the unadjusted standard-month equivalent; this polynomial reproduces the tabulated values of the original scheme, in which PET above the threshold depends only on temperature [7]. Positive-temperature conditions prevailed during the study period, and the same thermal function was used in the monthly and daily calculations to preserve **comparability.**

The daily adaptation followed Pereira and Pruitt [11]. Effective temperature $\left(T_{ef}\right)$ was calculated using daily mean temperature $\left(T_{avg}\right)$, daily thermal amplitude $A = Tmax - Tmin$, *and* $k = 0.69$, followed by a photoperiod adjustment. Because Tavg was computed as (Tmax + Tmin)/2, Equation (15) is algebraically identical to the original form Tef = 0.5k(3Tmax − Tmin) [11]:

$$T_{ef} = 0.69\left(T_{avg} + A\right) \quad (15)$$

$$T_{ef^*} = T_{ef} \frac{N}{(24 - N)} \quad (16)$$

where N is daylight duration (h); the adjusted effective temperature was constrained to $Tavg \leq Tef* \leq Tmax$. Daily PET was obtained from the corresponding Thornthwaite thermal function multiplied by $N/360$, and monthly totals were calculated by summing daily values. The coefficient k = 0.69, the photoperiod factor N/(24 − N) and the bounds Tavg ≤ Tef* ≤ Tmax were taken directly from [11]; the application-specific choices were the 2019–2025 thermal index I and the high-temperature polynomial, both applied identically in the monthly and daily calculations. Near the equator N ≈ 12 h, so N/(24 − N) ≈ 1 and, from Equation (15), Tef* ≈ k(Tavg + A). This value falls below Tavg whenever A < Tavg(1/k − 1) ≈ 0.45 Tavg, that is, whenever the daily thermal amplitude is

smaller than about 12 °C for the mean Tavg of the station (26.9 °C); the mean amplitude was 9.3 °C. Under the humid conditions of the station this condition is expected on most days, so the lower bound is expected to be active and the adaptation then reduces to the Thornthwaite thermal function evaluated with daily Tavg. The literature-derived coefficient was not calibrated locally; the adaptation was therefore evaluated as an alternative daily formulation, not as an independently validated model of measured ETo. The frequency of activation of each bound was recorded, and the contribution of the lower bound to the difference between the daily and monthly formulations was quantified by recomputing daily PET without the bounds (Table S8).

2.10. Statistical Performance Assessment

Agreement between the four alternative methods and FAO-56 PM was evaluated at daily and monthly scales using MAE, RMSE, MBE, percentage bias (PBIAS), Pearson's r, predictive R^2 (equivalent to the Nash–Sutcliffe efficiency, NSE), and Willmott's index of agreement (d) [33,34]. MAE, RMSE, and MBE followed Equations (1)–(3), with P_i now representing the estimate of an alternative ETo method and O_i the corresponding PM estimate. Additional measures were calculated as:

$$PBIAS = 100\frac{\sum(P_i - O_i)}{\sum O_i} \tag{17}$$

$$R^2 = 1 - \frac{\sum(P_i - O_i)^2}{\sum\left(O_i - \overline{O}\right)^2} \tag{18}$$

$$d = 1 - \frac{\sum(P_i - O_i)^2}{\sum\left(\left|P_i - \overline{O}\right| + \left|O_i - \overline{O}\right|\right)^2} \tag{19}$$

where $\bar{O}$ is the mean PM estimate. With this denominator, the predictive R^2 is algebraically identical to the Nash–Sutcliffe efficiency [35] computed with PM as the reference; it differs from the square of Pearson's r and is negative when the squared errors exceed those of a constant prediction equal to the PM mean. The denominator of Equation (18) was printed incorrectly in the original submission as $\Sigma(P_i - \bar{O})^2$; the reported values were computed with the conventional definition, as confirmed by an independent recomputation that reproduced every published value to the third decimal and by the identity $1 - \mathrm{RMSE}^2/\sigma^2(\mathrm{PM})$. The indicators quantify agreement with an equation-based reference and do not constitute validation against independently measured evapotranspiration. For the daily and monthly analyses, the evaluated pairs were aligned on coincident dates. Positive MBE and PBIAS denote estimates exceeding PM, whereas negative values indicate lower estimates. Pearson's r describes linear association but cannot by itself establish agreement in magnitude; thus, bias and magnitude-based metrics were considered together.

The main daily comparison was based on 2713 full days and the standard monthly comparison on 89 full months (August 2018 to December 2025). We performed the same calculations for the QC, conservative and exploratory meteorological scenarios as a sensitivity assessment, and separately for the drier (December–March) and wetter (April–November) periods. Given that their temporal coverage was different we interpreted the scenario-level statistics in conjunction with the sample sizes and made comparisons on common dates where possible. This was necessary as the QC observations were concentrated in a relatively small portion of the period, whereas the full reconstruction allowed for continuous comparisons across the years with varying levels of measurement availability. Uncertainty in the performance metrics was quantified with a moving-block bootstrap [36], which resamples contiguous blocks of days or months and preserves temporal dependence. The daily block length was set to the lag at which the autocorrelation of the daily error series

first fell below 1/e, capped at 60 days. Blocks of 6 months were used for the monthly series. We drew 2000 replicates (500 for the breakdown by scenario and season) and reported 95% percentile intervals. Paired differences between methods (e.g., RMSE of PT minus RMSE of Thornthwaite) were computed within each bootstrap replicate on the same resampled dates. This ensures their intervals account for dependence between the two error series. A difference was considered supported when its 95% interval excluded zero. This procedure serves the same purpose as a test of equal predictive accuracy with autocorrelation-consistent variance [37]. Results by data-provenance scenario, season, and validation block are reported in the Supplementary Materials (Tables S3 and S7).

2.11. Seasonal, Meteorological, and ENSO Analyses

We assessed seasonal variation using calendar-month means from 2019 to 2025. We calculated annual ET sums for the seven complete years. We evaluated differences from PM at daily, monthly, and annual scales, including relative annual differences. We analyzed relationships between daily PM and mean air temperature, RH, global solar radiation, wind speed, vapor pressure deficit, and net radiation using Pearson's r for all complete days and, separately, for the fully QC-observed days. Because several of these meteorological variables are inputs to PM, we interpreted these relationships descriptively rather than causally. Monthly anomalies were computed relative to the corresponding calendar-month mean over 2019–2025 and standardized using the standard deviation of the anomaly series. ENSO variability was characterized from NOAA Climate Prediction Center Relative Oceanic Niño Index (RONI) values [38], based on ERSST version 6 (snapshot of 20 October 2026); the monthly series, the identified episodes and the phase assigned to each month are listed in Table S9. The index's overlapping three-month seasons were assigned to their central calendar month. El Niño and La Niña episodes were identified by RONI thresholds of +0.5 and −0.5 °C, respectively, persisting for at least five consecutive overlapping seasons; the rule was applied to the full RONI record rather than only to the study period, so that episodes beginning before August 2018 or ending after December 2025 were classified consistently. Remaining months were treated as neutral. Because the 89 months are serially dependent, belong to few episodes and are strongly unbalanced in observational support, they do not constitute independent ENSO evaluations; no inferential test was attempted, and this analysis is reported as exploratory. Monthly PM estimates and anomalies were compared across ENSO phases together with the fraction of days with fully observed and reanalysis-assisted inputs in each phase, because both the ETo estimates and the meteorological anomalies partially reflected reanalysis-assisted data reconstruction.

3. Results

3.1. Meteorological Data Quality and Reconstruction Performance

The examination of the original La Libertad station records identified substantial interruptions and physically inconsistent measurements before evapotranspiration estimation. The initial database contained 50,805 hourly records, compared with 65,120 timestamps in the continuous index established for 28 July 2018–31 December 2025. Among the anomalies detected were 9110 wind-speed readings equal to or above 30 m/s, 1296 relative-humidity values exceeding 100%, 641 negative global-radiation readings, and 1977 zero values for minimum air temperature. The wind-speed anomalies were especially concentrated in periods when most other variables were unavailable. Cross-sensor examination additionally identified persistent irregularities in global and visible radiation, requiring extended quarantine intervals rather than interpolation between suspect measurements.

After QC, only 527 days had complete, accepted hourly observations for all six variables (Table 2), and their temporal distribution was very uneven: 61 days in the partial year

2018, 348 in 2019, 110 in 2020, none in 2021–2024 and eight in 2025. The conservative and exploratory scenarios increased daily coverage to 950 and 1511 days, respectively, whereas the complete multistage RF–ERA5 dataset had 2713 full daily records. The complete dataset comprised 527 fully observed days, 423 days with short-gap RF estimates, 528 days with long-gap exploratory estimates and 1235 days with hybrid RF–ERA5 predictions. The exploratory scenario comprises the first three categories (1478 days) plus 33 of the RF–ERA5 days, whose short gaps had been filled with temporal alternatives in the exploratory dataset (Section 2.4). These categories refer to the most reconstruction-dependent input of each day rather than to the replacement of every hourly observation.

Table 2. Annual availability of complete days by data-provenance category and duration of the radiation quarantine.

Year	Complete Days	Fully QC	Short-Gap RF	Long-Gap RF	RF–ERA5	Fully QC (%)	Rs Quarantine (Days)
2018 [a]	156	61	84	9	2	39.1	0
2019	365	348	5	0	12	95.3	0
2020	366	110	0	0	256	30.1	0
2021	365	0	0	73	292	0.0	74
2022	365	0	0	288	77	0.0	365
2023	365	0	0	14	351	0.0	365
2024	366	0	45	144	177	0.0	321
2025	365	8	289	0	68	2.2	0
Total	2713	527	423	528	1235	19.4	1125

Days are classified by the most reconstruction-dependent input within each day. [a] 29 July–31 December 2018 (the first, incomplete day was excluded). Rs quarantine: 19 October 2021–16 November 2024, inclusive. Cumulative scenarios (Section 2.4): observed QC, 527 days; conservative (QC + short-gap RF), 950; exploratory, 1511 = 527 + 423 + 528 + 33, where the 33 days (6 in 2019 and 27 in 2025) had short gaps filled with temporal alternatives in the exploratory dataset and are counted in the RF–ERA5 column because those hours were filled with hybrid RF–ERA5 predictions in the final series; complete, 2713.

The 2021–2024 period warrants particular caution. During these years, each complete day included at least one reconstructed variable, and radiation was fully reconstructed from 19 October 2021 to 16 November 2024 (Table 2). Despite this, station observations were QC-accepted for 37.3% of hourly Tmin values, 42.6% of Tmean and Tmax, 49.0% of RH, 53.7% of wind speed, and 14.1% of radiation values in 2021–2024. Consequently, the temperature, humidity, and wind series for this period are partially observed, while radiation is almost entirely based on reanalysis. The meteorological characteristics for this period are therefore estimates conditioned on ERA5-Seamless and on relationships derived from 2018–2021 and 2024–2025 observations. Reliability can only be assessed indirectly through blocked and leave-one-year-out validation (Tables 3 and S4), preservation of inter-variable correlations (Table S5), and propagation of input errors to ETo (Table S6), as reported below. Therefore, results for 2021–2024 should be interpreted as reanalysis-assisted estimates rather than direct climate observations.

Blocked validation showed that the hybrid RF–ERA5 reconstruction outperformed all simpler alternatives for every variable (Table 3). Relative to the calendar-only reference, RMSE decreased from 2.00 to 1.71 °C for Tmin, from 2.07 to 1.77 °C for Tmean, and from 2.13 to 1.81 °C for Tmax (14–15%). It also decreased from 9.45 to 8.04 percentage points for RH (14.9%), from 135.9 to 121.4 W/m^2 for global radiation (10.7%), and from 1.08 to 1.02 m/s for wind speed (4.8%). Correlations for the temperature variables reached 0.87–0.88, compared with 0.82–0.84 for the calendar reference. Raw ERA5-Seamless showed systematic biases at the station (MBE of −1.73 °C for Tmax, −7.2 percentage points for RH, +18.5 W/m^2 for radiation, and −1.25 m/s, or −54%, for wind speed after reduction to 2 m). The hybrid

model reduced RMSE relative to raw reanalysis by 10% (radiation) to 42% (wind speed). The monthly linear bias correction removed most of the mean bias but improved on the calendar reference only for radiation. The hybrid model still reduced RMSE relative to it by 7% (radiation) to 21% (RH). Hybrid MBE values were small (≤0.02 °C for temperature, +1.0 percentage point for RH, −8.7 W/m^2 for radiation, and +0.10 m/s for wind speed), indicating that the improvement did not come at the cost of systematic bias. Performance varied among blocks (RMSE ranges in Table 3; block-level results in Table S3). The hybrid model was not superior in every block; for example, in the 7-day block of 2018, its RMSE for Tmax (2.03 °C) exceeded that of the calendar reference (1.88 °C). Validation statistics were calculated only at withheld timestamps with QC observations (from 4240 for solar radiation to 5089 for mean and maximum temperatures and wind speed, owing to variable-specific QC exclusions within the blocks). Therefore, they should not be interpreted as verification of every reconstructed hour.

Table 3. Blocked validation of hourly meteorological reconstruction: calendar-only RF, raw ERA5-Seamless, linearly bias-corrected ERA5-Seamless, and hybrid RF–ERA5.

Variable/Method	n	MAE	RMSE	MBE	r	RMSE Range Across Blocks	ΔRMSE vs. Calendar (%)
Tmin (°C)							
Calendar-only RF	4994	1.44	2.00	+0.28	0.82	1.69–2.23	—
Raw ERA5 [a]	4994	1.62	2.10	−0.66	0.82	1.67–2.43	−4.9
Bias-corrected ERA5 [b]	4994	1.55	2.04	−0.06	0.80	1.67–2.19	−1.6
Hybrid RF–ERA5	4994	1.23	1.71	−0.02	0.87	1.34–1.91	+14.7
Tmean (°C)							
Calendar-only RF	5089	1.48	2.07	+0.28	0.83	1.71–2.33	—
Raw ERA5 [a]	5089	1.85	2.35	−1.18	0.83	1.73–2.79	−13.7
Bias-corrected ERA5 [b]	5089	1.63	2.13	−0.13	0.81	1.67–2.34	−2.8
Hybrid RF–ERA5	5089	1.27	1.77	+0.00	0.87	1.32–2.01	+14.3
Tmax (°C)							
Calendar-only RF	5089	1.53	2.13	+0.28	0.84	1.72–2.42	—
Raw ERA5 [a]	5089	2.18	2.73	−1.73	0.84	1.94–3.36	−28.4
Bias-corrected ERA5 [b]	5089	1.71	2.23	−0.16	0.82	1.72–2.49	−4.9
Hybrid RF–ERA5	5089	1.29	1.81	+0.01	0.88	1.32–2.05	+14.8
RH (percentage points)							
Calendar-only RF	5037	7.16	9.45	+0.71	0.76	6.82–10.57	—
Raw ERA5 [a]	5037	10.76	13.25	−7.18	0.72	10.14–14.60	−40.3
Bias-corrected ERA5 [b]	5037	7.97	10.15	+0.78	0.72	8.01–12.55	−7.4
Hybrid RF–ERA5	5037	5.90	8.04	+1.02	0.83	5.87–10.40	+14.9
Global radiation (W m^{-2})							
Calendar-only RF	4240	72.1	135.9	+4.1	0.88	121.2–147.4	—
Raw ERA5 [a]	4240	68.6	135.3	+18.5	0.89	106.1–144.5	+0.5
Bias-corrected ERA5 [b]	4240	69.2	130.9	−9.1	0.89	105.4–149.8	+3.7
Hybrid RF–ERA5	4240	65.6	121.4	−8.7	0.90	98.8–135.5	+10.7
Wind speed (m s^{-1})							
Calendar-only RF	5089	0.76	1.08	+0.10	0.55	1.03–1.14	—
Raw ERA5 [a]	5089	1.33	1.77	−1.25	0.24	1.55–1.88	−64.1
Bias-corrected ERA5 [b]	5089	0.96	1.23	+0.05	0.25	1.08–1.39	−14.7
Hybrid RF–ERA5	5089	0.73	1.02	+0.10	0.60	0.87–1.07	+4.8

Metrics computed at withheld hourly timestamps with QC observations in six blocks of 7–90 days (5208 h), from an independent re-run in which all steps were refitted within folds. MAE, RMSE and MBE are in the units of each variable. [a] Wind speed reduced from 10 m to 2 m with the FAO-56 logarithmic profile. [b] Monthly ordinary least-squares correction fitted outside each block. ΔRMSE: positive values indicate lower RMSE than the calendar-only reference.

In the leave-one-year-out experiment, the hybrid RMSE for the excluded years 2019 and 2020 was 1.92 and 1.75 °C for Tmax, 8.92 and 9.18 percentage points for RH, 111.6 and 100.7 W/m^2 for radiation, and 1.01 and 0.91 m/s for wind speed. These results are comparable to the pooled block results, and the hybrid model outperformed the calendar

reference for all variables (Table S4). In the 120- and 180-day blocks, RMSE increased moderately compared to the pooled 7–90-day blocks (Tmax 2.00 and 2.04 °C; RH 9.44 and 9.52 percentage points; radiation 119.9 and 120.8 W/m^2; wind speed 1.04 and 1.05 m/s), representing increases of up to approximately 18% for RH and 13% for Tmax. Removing the calendar-year predictor increased RMSE in both excluded years (e.g., RH 9.74 and 11.24 percentage points), indicating that the predictor did not degrade extrapolation and was therefore retained. The correlation structure of the daily inputs was only partly preserved (Table S5). Relationships among Tmax, RH and Rs were reproduced closely ($|\Delta r| \leq 0.10$; e.g., Tmax–Rs 0.84 observed vs. 0.87 reconstructed and Tmax–RH −0.74 vs. −0.85), whereas those involving Tmin were not (Tmax–Tmin 0.38 vs. −0.01; Tmin–Rs 0.18 vs. −0.25; Tmin–RH −0.44 vs. −0.13); the mean absolute difference across the ten pairs was 0.17. Consequently, the reconstructed daily thermal amplitude, which enters HS and the net longwave term, is less reliable than the other inputs. When ETo was computed for the 114 withheld days with complete observations from reconstructed rather than observed inputs, PM ETo showed RMSE = 0.97 mm/day, MBE = −0.18 mm/day and r = 0.54 (Table S6). Individual daily ETo values in reconstructed periods therefore carry errors of about 1 mm/day, comparable to the differences among methods, although these errors are partly shared by PM and the alternative methods because they use the same inputs. The ranking of the methods relative to PM was nevertheless identical with observed and reconstructed inputs (PT, Thornthwaite–Pereira–Pruitt, HS, BC), with PT RMSE values of 0.49 and 0.45 mm/day, respectively.

3.2. Reference Evapotranspiration Estimates and Daily Method Agreement

The final meteorological dataset analysis showed that the FAO-56 Penman–Monteith (PM) method gave a mean daily ETo of 3.78 mm/day (SD = 1.01 mm/day; range 1.03 to 7.70 mm/day). The radiative and aerodynamic components contributed on average 2.67 and 1.11 mm/day respectively. None of the 2713 PM values calculated were negative. All of the alternative methods produced very different values in terms of both magnitude and day-to-day variability (Table 4). The PT method averaged 4.02 mm/day (SD = 0.96), the HS method 4.57 mm/day (SD = 0.69), the daily Thornthwaite–Pereira–Pruitt 4.55 mm/day (SD = 0.56), and the daily Blaney–Criddle 5.45 mm/day (SD = 0.15). The BC method produced by far the narrowest distribution of values (despite having the highest mean). For the 527 days which were fully QC observed the mean PM ETo was 4.23 mm/day. The corresponding statistics for all of the methods are given in Table 4.

Table 4. Descriptive statistics of daily evapotranspiration estimates (mm/day) for all complete days and for the subset of fully QC-observed days.

Method	Subset	n	Mean	SD	Minimum	Maximum
FAO-56 PM	All complete days	2713	3.78	1.01	1.03	7.70
	Fully QC-observed	527	4.23	1.21	1.11	7.70
Hargreaves–Samani	All complete days	2713	4.57	0.69	2.05	6.72
	Fully QC-observed	527	5.14	0.65	2.82	6.72
Priestley–Taylor	All complete days	2713	4.02	0.96	1.10	6.72
	Fully QC-observed	527	4.19	1.02	1.12	6.54
Blaney–Criddle (daily)	All complete days	2713	5.45	0.15	4.65	6.03
	Fully QC-observed	527	5.53	0.19	4.97	6.03
Thornthwaite–Pereira–Pruitt	All complete days	2713	4.55	0.56	1.84	5.80
	Fully QC-observed	527	4.91	0.54	2.40	5.80

All values in mm/day. "All complete days" comprises 527 fully QC-observed days and 2186 days with at least one reconstructed input (423 short-gap RF, 528 long-gap RF and 1235 RF–ERA5 days).

The scatter diagrams of the daily alternative estimates against PM (Figure 2) show that PT followed the reference most closely, with reasonably good agreement across most of the observed range. HS estimates tended to be above the 1:1 line, particularly at the lower PM values, whereas BC estimates clustered within a narrow band of about 5–6 mm/day irrespective of the reference ETo. The daily Thornthwaite adaptation showed more temporal variability than BC but tended to be above PM on the low demand days and below the highest PM values. The patterns were broadly similar for fully QC-observed days and days with reconstructed inputs, although on fully QC-observed days PT fell below PM at the highest reference values (PM > 5.5 mm/day). These patterns are consistent with the different meteorological inputs and empirical assumptions of the methods; because no independent evapotranspiration measurements were available, they cannot be interpreted as errors relative to measured ETo.

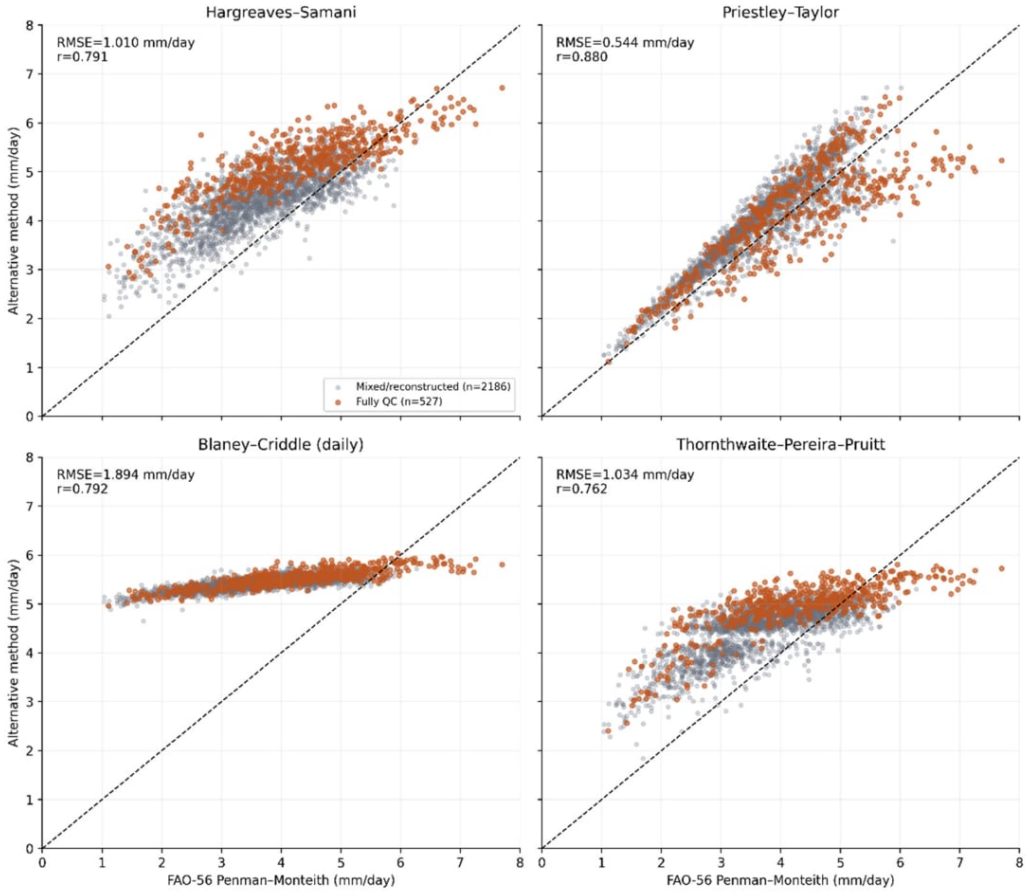


Figure 2. Daily agreement of the Hargreaves–Samani, Priestley–Taylor, daily Blaney–Criddle and Thornthwaite–Pereira–Pruitt estimates with FAO-56 Penman–Monteith (n = 2713 days). Orange circles: days with all six variables fully observed and QC-accepted (n = 527); gray circles: days with at least one reconstructed input (n = 2186); dashed line: 1:1 line. Statistics in each panel refer to all days.

The performance metrics confirm the graphical results (Table 5; Figure 3). PT had the lowest daily RMSE (0.54 mm/day; 95% CI 0.49–0.60) and MAE (0.46 mm/day) with an MBE of +0.25 mm/day (PBIAS = +6.5%), r = 0.88, NSE = 0.71 and d = 0.92. HS (RMSE = 1.01 mm/day; PBIAS = +20.9%) and the daily Thornthwaite adaptation (RMSE = 1.03 mm/day; PBIAS = +20.4%) were similar in terms of performance and considerably worse. Their paired RMSE difference was −0.02 mm/day (95% CI −0.07 to 0.02) so the two methods were not distinguishable from one another. BC had the poorest agreement

(RMSE = 1.89 mm/day; PBIAS = +44.1%; NSE = −2.49). The negative NSE indicates that the sum of the squared deviations was actually greater than that of using the mean PM value as a constant. Despite this, there was a positive correlation (r = 0.79). The daily advantage of PT over the other three methods was again confirmed by the paired bootstrap intervals which excluded zero (PT − HS = −0.47 mm/day, 95% CI −0.56 to −0.38; PT − Thornthwaite–Pereira–Pruitt = −0.49 mm/day, 95% CI −0.57 to −0.41). The ranking was not simply a result of the mean bias. PT also did a better job at reproducing the spread of the daily PM estimates, whereas the very low variance of BC meant it was not capable of distinguishing between days of low and high evaporative demand. Hence the positive correlations for the temperature-based methods did not necessarily equate to good agreement with the PM reference.

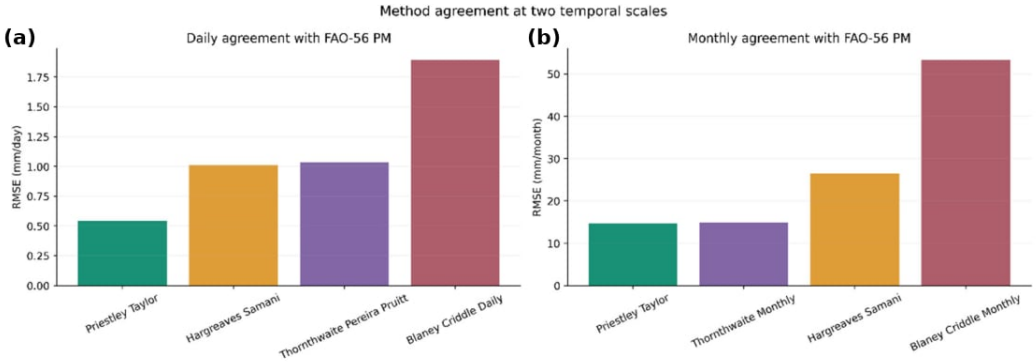


Figure 3. Root mean square errors of the four alternative methods relative to FAO-56 Penman–Monteith at (**a**) daily and (**b**) conventional monthly timescales. Methods were evaluated separately at each time resolution; 95% confidence intervals are given in Table 5.

Table 5. Agreement statistics relative to FAO-56 Penman–Monteith at daily (n = 2713 days) and conventional monthly (n = 89 months) timescales.

Scale	Method	n	Mean	MBE	MAE	RMSE (95% CI)	PBIAS (%)	r	NSE	d
Daily (mm d^{-1})	PM (reference)	2713	3.78	—	—	—	—	—	—	—
	HS	2713	4.57	+0.79	0.86	1.01 (0.94–1.08)	+20.9	0.79	0.01	0.73
	PT	2713	4.02	+0.25	0.46	0.54 (0.49–0.60)	+6.5	0.88	0.71	0.92
	BC daily	2713	5.45	+1.67	1.69	1.89 (1.79–2.01)	+44.1	0.79	−2.49	0.45
	T–Pereira–Pruitt	2713	4.55	+0.77	0.88	1.03 (0.97–1.09)	+20.4	0.76	−0.04	0.68
Monthly (mm $month^{-1}$)	PM (reference)	89	115.1	—	—	—	—	—	—	—
	HS	89	139.1	+24.0	24.1	26.5 (22.9–29.7)	+20.8	0.77	−1.29	0.61
	PT	89	122.6	+7.4	13.1	14.8 (12.5–16.2)	+6.5	0.71	0.29	0.79
	BC classical	89	165.8	+50.7	50.7	53.3 (48.7–58.9)	+44.1	0.36	−8.26	0.35
	T classical	89	125.7	+10.6	12.1	14.9 (12.0–17.6)	+9.2	0.80	0.28	0.80

Mean, MBE, MAE and RMSE are in mm d^{-1} (daily) or mm $month^{-1}$ (monthly); PBIAS in %; r, NSE (predictive R^2, Equation (18)) and d are dimensionless. 95% CI: moving-block bootstrap (Section 2.10). Paired RMSE differences (95% CI): PT − HS −0.47 (−0.56 to −0.38), PT − T–Pereira–Pruitt −0.49 (−0.57 to −0.41), HS − T–Pereira–Pruitt −0.02 (−0.07 to 0.02) mm d^{-1}; PT − T classical −0.1 (−4.3 to 3.1) mm $month^{-1}$. Results by scenario and season are given in Table S7.

3.3. Monthly Formulations, Seasonal Variability, and Annual Totals

Over the conventional monthly timescale (89 complete months from August 2018 to December 2025) the mean PM estimate was 115.1 mm/month (SD = 17.6). The corresponding mean values were 122.6 mm/month for PT, 139.1 mm/month for HS, 165.8 mm/month for classical BC, and 125.7 mm/month for classical Thornthwaite (Table 5). Monthly BC values were fairly stable (SD = 4.4 mm/month), ranging from 152.8 to 177.2 mm/month. PM, by contrast, ranged from 79.7 to 164.8 mm/month. The difference in temporal dispersion is clear in Figure 4, where BC is essentially flat, while PM, PT, and HS all show much greater variation.

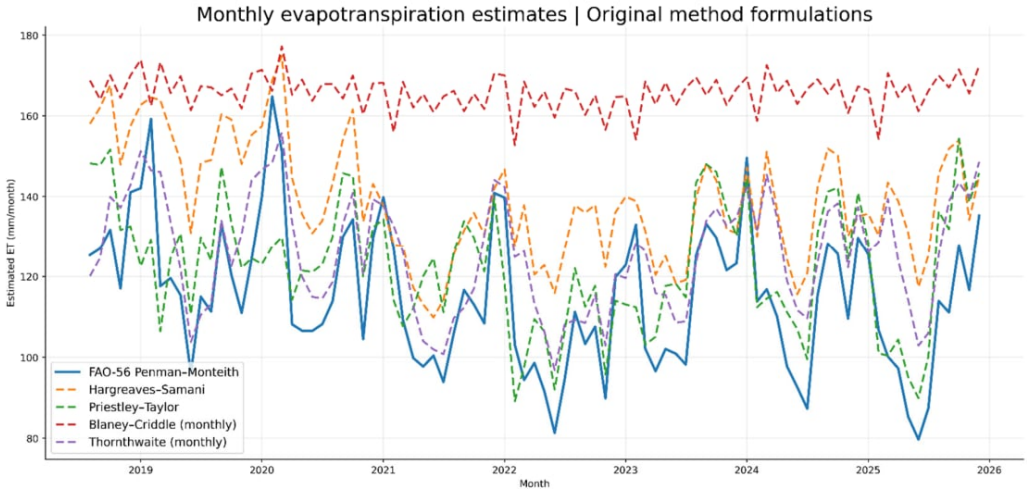


Figure 4. Monthly evapotranspiration estimates obtained using the conventional formulations of the five methods during August 2018–December 2025.

At this scale, PT and classical Thornthwaite showed similar agreement with PM. PT had a marginally lower RMSE (14.8 vs. 14.9 mm/month), but the paired bootstrap difference (−0.1 mm/month; 95% CI −4.3 to 3.1) included zero, and Thornthwaite had a lower MAE (12.1 vs. 13.1 mm/month), a higher correlation (r = 0.80 vs. 0.71) and a similar index of agreement (d = 0.80 vs. 0.79). The monthly ranking therefore depends on the criterion, and the data do not support the superiority of either method at this scale. Monthly HS and BC produced RMSE values of 26.5 and 53.3 mm/month, respectively, and monthly PBIAS values were +6.5% for PT, +9.2% for Thornthwaite, +20.8% for HS, and +44.1% for BC (Figure 3).

The daily BC extension produced monthly totals essentially identical to those from the conventional formulation (mean absolute difference of 0.002 mm/month over the full series). For Thornthwaite, the daily Pereira-Pruitt extension gave a mean monthly total of 138.5 mm (vs. 125.7 mm from the conventional method), an increase of about 10.2%. On 93.6% of complete days, the adjusted effective temperature reached its lower limit (Tef* = Tavg), and the upper limit was never reached, as expected from the analytical condition derived in Section 2.9.5. Recomputing daily PET without bounds yielded a mean monthly total of 111.2 mm (11.5% below the conventional formulation), so the active lower bound increased it by 27.3 mm/month. Conversely, if you evaluate the thermal function directly using daily Tavg, you get 138.2 mm/month, essentially the same as the bounded adaptation (Table S8). The 10% excess over the conventional formulation is therefore not due to the Pereira-Pruitt adjustment itself but mainly to the temperature used. The daily calculation uses Tavg = (Tmax + Tmin)/2, and its mean (26.9 °C) is about 1.0 °C higher than the mean of the hourly temperatures (25.9 °C) used in the monthly formulation. The thermal function

increases by about 7–13% per °C in this range, so using Tavg leads to an increase of about 10% in the monthly total. The high-temperature polynomial ($T \geq 26.5$ °C) was applied on 60.8% of the days. The two Thornthwaite formulations were therefore not interchangeable under the prevailing near-equatorial conditions. The mean seasonal PM series showed its highest daily values in February (4.59 mm/day) and January (4.42 mm/day), decreasing to 3.13–3.21 mm/day during May–July. A secondary increase occurred in September and October (4.08 and 3.96 mm/day), whereas the corresponding BC daily estimates varied much less between months (approximately 5.38–5.58 mm/day). Annual PM totals for the seven complete years ranged from 1235 mm in 2022 to 1498 mm in 2020 (Figure 5); PT followed the main interannual pattern, whereas conventional BC remained near 1959–2011 mm/year, and HS and conventional Thornthwaite showed intermediate magnitudes and differing annual variability. In 2022, for example, the annual PM total was 1235 mm, compared with 1282 mm for PT, 1568 mm for HS, and 1959 mm for conventional BC. Because 2021–2024 contained no fully QC-observed days and radiation was entirely reconstructed from October 2021 to November 2024 (Table 2), these interannual patterns describe the reanalysis-assisted series and should not be interpreted as observed interannual variability. In the partial year 2018, monthly analyses were restricted to August–December to avoid introducing incomplete monthly totals.

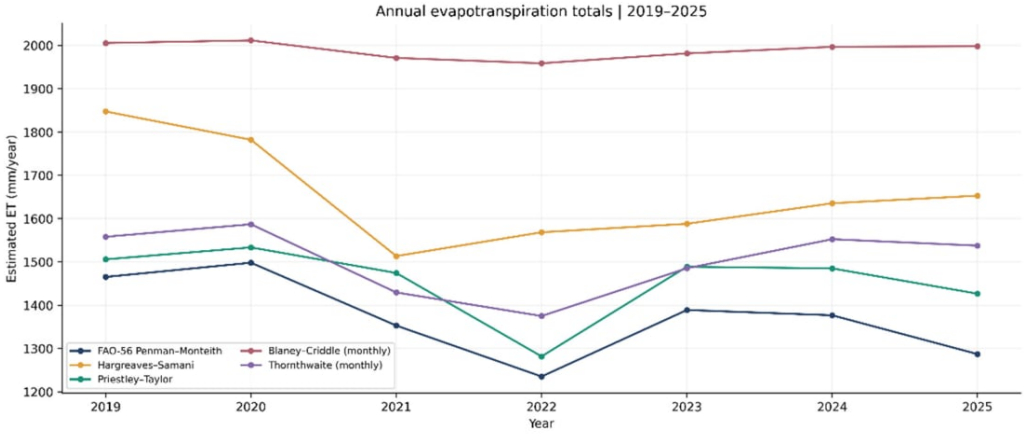


Figure 5. Annual evapotranspiration totals for the seven complete years (2019–2025), using the original monthly formulations of Blaney–Criddle and Thornthwaite.

3.4. Influence of Meteorological Data Quality and Calculation Choices

The mean daily PM estimate depended on the temporal subset retained in each reconstruction scenario: 4.23 mm/day for 527 fully QC days, 3.98 mm/day for 950 conservative days, 3.81 mm/day for 1511 exploratory days, and 3.78 mm/day for the complete record. However, PM estimates for the same 527 QC dates were identical in the QC and final datasets; accordingly, the difference of 0.46 mm/day between the fully observed subset and the complete series mainly reflects the addition of other dates and meteorological conditions rather than a modification of accepted measurements.

PT ranked first in all data-provenance scenarios and in both seasons (Table S7). On fully QC-observed days, PT had RMSE = 0.68 mm/day (95% CI 0.52–0.84), PBIAS = −1.0% and d = 0.90; over the complete series, the corresponding values were 0.54 mm/day, +6.5% and 0.92. HS and the daily Thornthwaite adaptation alternated between second and third place, consistent with their very similar RMSE, and BC was consistently last, with the largest positive bias. The lower RMSE over the complete series does not imply that the reconstructed data are more accurate than the QC observations, because the two subsets differ in coverage and meteorological conditions. A same-date comparison of ETo computed

from observed and reconstructed inputs (Section 3.1; Table S6) yielded the same ranking. Seasonally, PT remained close to PM in both periods (RMSE = 0.57 mm/day in December–March and 0.53 mm/day in April–November), although its bias shifted from −5.8% in the drier period to +13.4% in the wetter period. The PM sensitivity assessment showed that alternative radiation and temperature conventions had little effect on the mean compared with the differences among estimation methods. Setting negative outgoing longwave-radiation estimates to zero changed the complete-series PM mean by −0.002 mm/day, and replacing the FAO-56 daily temperature with the mean hourly temperature changed it by −0.010 mm/day. The largest daily differences were 0.22 and 0.12 mm/day, respectively. These findings support the standard PM configuration for the main method comparison, while showing that daily differences in specific cases can exceed the average effects.

3.5. Meteorological Relationships and Monthly Anomalies

Daily PM estimates showed the strongest covariation with radiation and atmospheric-moisture variables (Figure 6). Global solar radiation had the highest Pearson correlation with PM (r = 0.89), followed by net radiation (r = 0.86), mean air temperature (r = 0.83), and vapor pressure deficit (r = 0.82). Relative humidity was inversely correlated (r = −0.82), while wind speed had a weaker positive correlation (r = 0.33). Similar patterns were found for the 527 fully observed QC days (r = 0.84 for solar radiation, 0.82 for temperature, 0.84 for vapor pressure deficit, −0.84 for relative humidity, and 0.43 for wind speed). Because these variables are direct inputs to the PM equation, the correlations reflect the equation's response to its inputs under local conditions. These results are descriptive and do not demonstrate independent causal control of evapotranspiration by any of these variables.

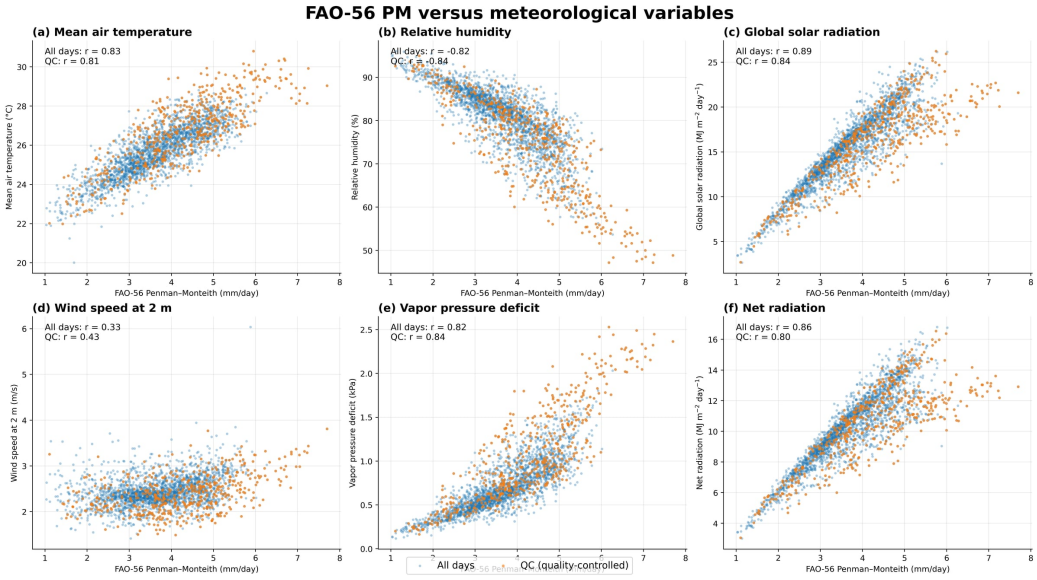


Figure 6. Daily FAO-56 Penman–Monteith estimates as a function of (**a**) mean air temperature, (**b**) relative humidity, (**c**) global solar radiation, (**d**) wind speed, (**e**) vapor pressure deficit, and (**f**) net radiation.

The monthly anomaly series (Figure 7) showed broadly similar behaviour in PM and a number of the other atmospheric variables, particularly solar radiation and humidity. The positive PM anomalies were generally associated with higher temperature/radiation and lower relative humidity, whilst the negative PM anomalies were generally associated with lower radiation and more humid conditions. The wind speed anomalies were less clearly correlated with the PM anomalies. Some of the co-variance is of course a mathematical consequence of these being the inputs to PM. The largest deviations were in the periods

for which we had the reconstructed meteorological record so the correspondence in the graph is a reflection of a property of the reconstructed meteorological record rather than independent evidence of the observed atmospheric variability.

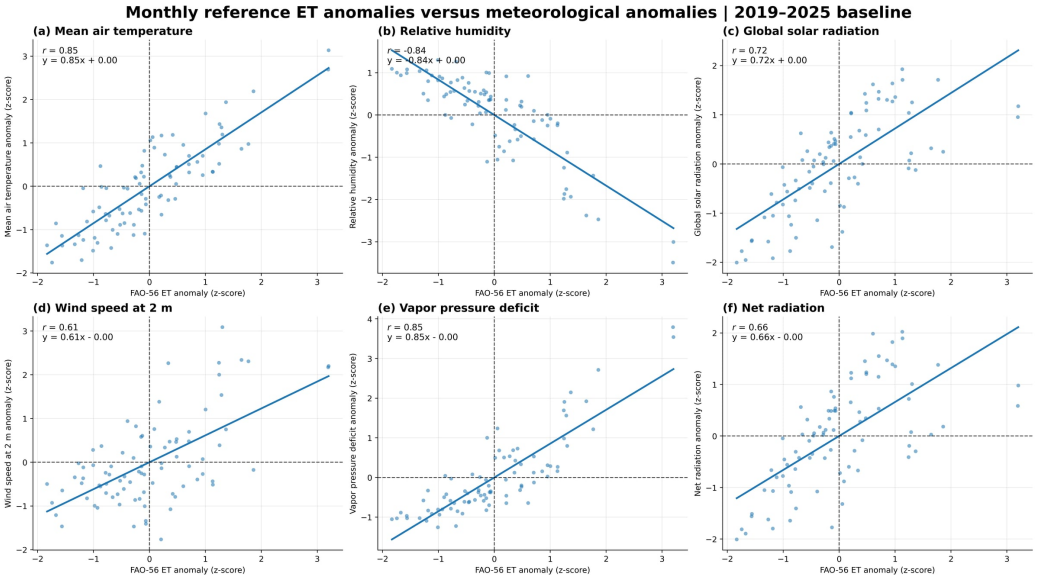


Figure 7. Standardized monthly anomalies of FAO-56 Penman–Monteith and selected meteorological variables relative to the 2019–2025 calendar-month baseline.

3.6. Exploratory Description of Reference Evapotranspiration Across ENSO Phases

We used the NOAA RONI to classify the 89 months as 15 El Niño, 23 neutral and 51 La Niña months (see Table S9 and Figure S2 for the monthly series). The average PM ETo was 4.22 mm/day in the El Niño months, 3.75 mm/day in the neutral months and 3.67 mm/day in the La Niña months. Relative to the 2019–2025 calendar month reference the mean anomalies were +0.26, +0.17 and −0.12 mm/day respectively (Table 6). The anomaly distributions overlapped considerably (Figure S3) so there was a lot of variability within each phase.

Table 6. Exploratory summary of monthly reference evapotranspiration by ENSO phase, with observational support and reconstruction fraction.

ENSO Phase	Months	Days	Mean PM (mm d^{-1})	Mean Anomaly (mm d^{-1})	Fully QC Days	Fully QC (%)	Days with Reconstructed Inputs (%)	Days with RF–ERA5 Inputs (%)
El Niño	15	456	4.22	+0.26	162	35.5	64.5	54.2
Neutral	23	701	3.75	+0.17	357	50.9	49.1	35.8
La Niña	51	1553	3.67	−0.12	8	0.5	99.5	47.5

Exploratory and descriptive; months are serially dependent and not independent ENSO evaluations. Phase classification from NOAA CPC RONI (ERSST version 6; snapshot of 20 October 2026), with the five-season rule applied to the full RONI record (Table S9). Days from August 2018 to December 2025 (n = 2710).

Observational support varied significantly by phase: only 0.5% of La Niña days were fully QC-observed (8 of 1553), compared with 35.5% of El Niño days (162 of 456) and 50.9% of neutral days (357 of 701). Most La Niña days (99.5%) relied on reconstructed inputs, including 32.9% with long-gap RF and 47.5% with hybrid RF–ERA5 estimates (Table 6). Phase, data availability, and the 2021–2024 radiation quarantine are strongly confounded, as the main La Niña episode (May 2020–March 2023) aligns with the reanalysis-dominated period. Additionally, the 89 months are serially dependent and represent only two El Niño (2018–2019 and 2023–2024) and three La Niña episodes (2020–2023, 2024–2025, and from

August 2025), so they do not provide 89 independent ENSO evaluations. We therefore did not conduct an inferential test and report phase differences only as a descriptive feature of the reconstructed 2018–2025 series.

4. Discussion

In this section, "agreement" and "error" refer to differences relative to FAO-56 PM calculated from the same (mostly reconstructed) meteorological inputs. Because no independent evapotranspiration measurements (lysimeter or eddy covariance) were available, the analysis evaluates how closely the simplified formulations track the standard equation, not their accuracy in estimating actual or reference evapotranspiration.

4.1. Meteorological Reconstruction as a Prerequisite for ETo Assessment

The first implication of this study concerns the meteorological record rather than the evapotranspiration equations themselves. Although the La Libertad station provided more than 50,000 original hourly records, strict quality control identified long intervals during which a conventional complete-case analysis could not support a continuous multi-year comparison. Only 527 of the 2713 complete daily records were based entirely on accepted observations, with no fully QC-observed days during 2021–2024. Thus, simply applying FAO-56 Penman–Monteith (PM) to the unprocessed station file would have produced a temporally selective picture of atmospheric evaporative demand, recovering coverage required to distinguish isolated sensor gaps, for which other station variables remained informative, from prolonged outages requiring external information.

This distinction aligns with the problem identified by [15] in southwestern Colombia, which used records from 1983–2017; they compared empirical ETo equations and machine-learning approaches with limited meteorological data. Their best monthly ARIMA configuration achieved RMSE = 4.13 mm/month and r = 0.98 relative to FAO-56 PM, but the authors warned that a locally fitted model with exogenous variables could not be directly transferred to another station. This work had a different objective: rather than predicting ETo directly, we reconstructed the meteorological inputs before applying five independently defined equations; this preserves the possibility of comparing their physical and empirical assumptions on a common dataset, although it also propagates meteorological reconstruction uncertainty to every subsequent estimate.

Likewise, ref. [19] showed that machine-learning and deep-learning approaches can estimate daily ETo from limited meteorological information, with temporal convolutional networks outperforming the empirical methods examined in their northeastern China application. More recently, ref. [10] evaluated several regression and neural-network approaches using data from Colombia's Cauca River Valley; they found that predictor combinations including radiation and wind were more informative than combinations dominated by temperature and humidity; some configurations reached R^2 values near 0.98 against PM. These studies support the usefulness of meteorological predictors and nonlinear models, but neither establishes the accuracy of a reconstructed station variable during a multi-year instrument outage. In this case, the hybrid procedure improved on both calendar-based estimates and raw reanalysis (Table 3), with the smallest gain for wind; this improvement is meaningful, but it reflects agreement with withheld QC observations, not proof that every filled hour reproduces the unobserved local weather.

Applying reanalysis data requires careful consideration of scale. For example, ref. [30] showed that ERA5-Land offers hourly land-surface fields at about 9-km resolution and generally improves water-cycle representation compared to ERA5. However, since reanalysis is a modeled representation over a grid cell, the product was evaluated against in situ observations. At La Libertad, ERA5-Seamless provided valuable continuous predictors, but

grid-cell values should not be treated as direct measurements of local wind, humidity, or radiation. This distinction is especially important at the Andean–Orinoquian transition, where terrain, convection, and cloud cover can vary at scales smaller than the reanalysis grid (Section 2.5; Figure 1a). The separate provenance flags used in this study are not just for data management; they define the evidential boundaries for subsequent climatic comparisons.

4.2. Why Priestley–Taylor Showed the Closest Agreement with Penman–Monteith

PT provided the best overall agreement with PM at La Libertad (Table 5). That makes physical sense in a humid tropical environment as PT explicitly incorporates the available energy and the equilibrium component of evaporation. In the PM calculation the radiative term was responsible for around 71% of the mean ETo and radiation had the best daily correlation with PM. The fact that PT performed so well in this case is in part structurally expected though as it was being used with the same net radiation as PM (including the humidity-dependent longwave term). The agreement is really just a measure of how much the aerodynamic term and the fixed coefficient $\alpha = 1.26$ differ from the PM aerodynamic term and should not be taken as any kind of independent validation.

Our results agree with [6], which compared five alternative daily ETo methods with FAO-56 PM in Garanhuns, northeastern Brazil, where they identified PT as the best-performing alternative by mean absolute error, while Hargreaves–Samani (HS), after calibration, was the best among the temperature-based approaches. These findings suggest that an energy-based equation may capture daily fluctuations more effectively when radiation is available. Ref. [39] provided a complementary, experimentally grounded interpretation by comparing PT-type formulations against weighing-lysimeter measurements from humid tropical Piracicaba, Brazil, and semi-arid Davis, California, and found that the inferred PT coefficient was approximately 1.20 in the tropical environment and 1.27 in the semi-arid environment, leading the author to propose a variable coefficient related to surface–atmosphere coupling. The PT bias of about +6.5% found here with the conventional coefficient of 1.26 is compatible with the possibility that a locally appropriate coefficient may differ, but our PM-based comparison cannot determine its lysimetrically optimal value.

Not every Colombian study has identified PT as the most accurate equation. Ref [3] evaluated ETo methods using CENICAÑA station data from 1994 to 2014 in the Cauca River Valley. In their comparison, PT showed a low correlation with PM ($r = 0.33$) and RMSE = 0.80 mm/day, whereas we obtained $r = 0.88$ and RMSE = 0.54 mm/day. Their mean PT value was 4.12 mm/day compared with 4.47 mm/day for PM, indicating underestimation, whereas PT exceeded PM at La Libertad. Thus, even with the same equations, performance can vary considerably with the radiation regime, atmospheric coupling, instrument setup and study period. Our results support PT as an uncalibrated alternative for this dataset, but not as the best method for Colombian tropical climates in general.

4.3. Limitations of Temperature- and Daylight-Based Methods

HS exceeded PM by about 21% on average and showed substantially weaker daily agreement than PT. Its reduced representation of short-term variability stems from relying on extraterrestrial radiation and the daily temperature range as a proxy for radiation reaching the surface. Although the astronomical term is known, actual incoming solar radiation is strongly modified by cloudiness, which is prominent in the Orinoquía. A similar concern motivated [40], who found that the uncalibrated Hargreaves equation generally overestimated ETo at humid locations in the western Balkans and investigated adjusting its exponent. His climatic context differs markedly from Villavicencio, but the direction of the bias agrees with our result and reinforces the need for site-specific calibration. By

contrast, ref. [3] reported a smaller HS excess at CENICAÑA, with averages of 5.06 and 4.47 mm/day for HS and PM, respectively.

The simplified Blaney–Criddle (BC) formulation has a significant limitation: its mean exceeds that of PM by 44%, and its daily standard deviation (0.15 mm/day) is less than one-sixth of PM's (1.01 mm/day). This pronounced bias and negative NSE indicate issues beyond a constant offset. The uncorrected temperature–daylight relationship does not capture the observed range of PM fluctuations. In equatorial regions, daylight duration is relatively constant year-round, and temperature alone does not account for rapid changes in humidity and radiation caused by cloud cover. Notably, ref. [3] found that BC was slightly below PM in mean magnitude at CENICAÑA (4.18 versus 4.47 mm/day), but with similarly limited variation (SD = 0.09 mm/day). This comparison underscores the difference between a method's ability to reproduce variability and the direction of its mean bias. It also demonstrates that applying a generic correction factor to a stable BC series would not necessarily restore the timing of high- and low-demand days. These conclusions apply only to the simplified expression evaluated in this study. Formulations with FAO-24 adjustments for humidity, sunshine, and wind [31] were not assessed and may perform differently. Therefore, these findings do not imply that all versions of the Blaney–Criddle method are inadequate.

The two Thornthwaite implementations highlight a different issue: temporal adaptation itself can materially affect the estimates. The classical monthly Thornthwaite was 9.2% above PM, whereas the daily Pereira-Pruitt was 20.4% above PM, and its monthly aggregates were about 10% higher than the conventional formulation. Ref. [11] tested their effective temperature and photoperiod adjustment against lysimeter data in two contrasting climates and found that the original formulation systematically underpredicted evapotranspiration in the arid Davis climate, whereas their proposed adaptation substantially reduced that systematic error (at the cost of a small overprediction). At La Libertad, the lower bound of the adjusted effective temperature was active on 93.6% of the days. As discussed in Section 2.9.5, this is exactly what you would expect given a near 12 h photoperiod and a daily thermal amplitude small relative to the mean temperature so adaptation was predominantly operating under an active constraint and effectively reduced the Thornthwaite thermal function to that evaluated with daily Tavg. The decomposition in Table S8 shows that the ~10% difference from the conventional formulation is largely due to using (Tmax + Tmin)/2, which is about 1 °C warmer than the mean of the hourly temperatures, rather than the effective temperature adaptation itself. Furthermore, the original Thornthwaite equation estimates potential evapotranspiration (PET), not FAO-56 reference evapotranspiration, so comparing it with PM compares two related but different concepts. Since our benchmark was PM, not lysimeter data, the higher daily Thornthwaite estimates cannot be interpreted as an improvement in physical accuracy, and both implementations should be treated as distinctly different when results are reported or used for water planning.

4.4. Temporal Scale, Meteorological Covariation, and Practical Implications

Aggregation had an effect on the apparent ranking of some of the equations. At the daily scale PT was the best performer with respect to PM (paired bootstrap intervals all excluding zero). At the monthly scale PT and the classical Thornthwaite were essentially the same. The 0.1 mm/month difference in RMSE is very small in the context of its sampling uncertainty (95% CI −4.3 to 3.1 mm/month) and the uncertainty in the reconstructed inputs. Thornthwaite had the lower MAE and the higher correlation. The results therefore indicate (i) better daily agreement for PT, (ii) similar performance on a monthly basis between PT and the conventional Thornthwaite and (iii) no evidence of physical equivalence or superiority of either method as both are being compared against an equation-based reference.

Monthly aggregation can allow daily deviations in opposite directions to cancel each other out and improve the cumulative agreement without necessarily restoring the daily sequence which is of relevance to irrigation scheduling. The very similar monthly totals produced by the conventional and daily BC formulations also confirm that a change in the timescale of computation does not in any way overcome the absence of the physical drivers. As [41] so rightly points out, simplified ETo approaches can be of value in the circumstances of poor or limited data, but apparent overall agreement in no way obviates the need to consider data quality and the application in question.

The global hydrological analysis of [42] provides a really useful perspective on what constitutes a 'best' method. Looking across 5338 streamflow gauges and independent remotely sensed ET data they found that the preferred PET formulations varied with climate and with the objective of the evaluation (whether it was ET or catchment discharge) and that Hargreaves performed well in many of the tropical catchments whereas PT was preferred in other circumstances and for some of the streamflow criteria; this may seem to contradict our own ranking but the objective is different. Our analysis is looking at equation-to-equation agreement for a hypothetical reference surface at one station; theirs is looking at the impact of PET choice on a calibrated catchment water balance and their conclusion was that the choice of PET has very significant consequences in the tropics which does support the idea that the positive biases we have observed here may be of relevance to irrigation and hydrological applications, without us necessarily having to assume that those percentages equate to actual crop water use or streamflow error.

Covariation with meteorological inputs adds descriptive context: PM is positively associated with global radiation, temperature, and vapor pressure deficit, and negatively with RH. Wind shows a weaker relationship (Section 3.5). This pattern aligns with higher reference evaporative demand on sunny, warm, and dry-air days, even in a generally humid climate. However, these correlations partly result from variables included in PM calculations and do not independently demonstrate process control over crop evapotranspiration. The next step would be to test calibrated equations using separate wet- and dry-season validation sets and, where possible, independent flux or lysimeter observations rather than relying solely on PM correspondence.

4.5. ENSO: An Exploratory Association Confounded by Data Provenance

Mean PM ETo was higher in the El Niño months than in the La Niña months (Table 6). This is broadly in line with the hydroclimatic studies which have reported warmer/drier conditions in parts of Colombia during El Niño and increased moisture availability during La Niña: [43] looked at the responses of precipitation, streamflow, soil moisture and vegetation to ENSO and found that the ENSO-induced hydrological anomalies were seasonally dependent with the response of streamflow being potentially amplified by soil moisture and ET processes. Ref. [44] also highlighted the interplay of atmospheric and topographic processes in shaping the ENSO signal in the Colombian Andes. Those studies provide a reasonable physical basis for the anomalies we are seeing, but do not in any way confirm that the same mechanisms are at work or that the same magnitude of effect is occurring at the Villavicencio station.

In the current record, ENSO phase is strongly confounded with data provenance. Almost all La Niña months occur in the reanalysis-dominated 2020–2023 period (only 8 fully QC-observed days), and the 89 months are serially dependent and clustered in a few episodes. Without controls for seasonality, temporal dependence, individual events and reconstruction fraction, phase differences cannot be separated from these effects and should not be interpreted as an ENSO impact on reference ETo. Longer records, independent

stations and appropriate controls are required to establish a robust ENSO-related response in the Orinoquía.

4.6. Methodological Limitations and Research Priorities

Our limitations to the interpretation of this comparison should be noted. Firstly, PM was the computational benchmark against which the results were assessed. Agreement with PM does not in any way validate the results against measured evaporation or crop evapotranspiration. As [3] points out, field-reference measurements are valuable when the empirical equations differ substantially. Second, we could validate the reconstruction only where QC observations were available. The blocked tests covered gaps of 7 to 90 days, and the leave-one-year-out and 120–180-day tests (Table S4) showed only modest degradation for the longer gaps. However, none of the tests replicated the full 37-month radiation quarantine, where radiation and the radiation-dependent PM and PT estimates relied entirely on the ERA 5-Seamless predictors. The reconstruction failed to preserve the covariation of Tmin with the other variables (Table S5). The PM ETo calculated from the reconstructed inputs on the validation days had an RMSE of 0.97 mm/day compared to the PM ETo from the observed inputs (Table S6). Thus, individual daily ETo values in the reconstructed periods should be treated as having an uncertainty of approximately 1 mm/day, although the relative ranking of the methods was unaffected. The reconstruction fraction was highest in 2021–2024 (all days had reconstructed inputs; Table 2) and in La Niña months (Table 6), so observations constrain the interannual and ENSO-related results the least. Thirdly, the fully observed subset is concentrated in 2018–2020 and is not a random sample of the period under study. Fourthly, the study is based on a single station. While this is appropriate for a local comparison, the results cannot be generalized to the Colombian Orinoquía or other humid tropical regions without a multi-station evaluation.

Finally, the fixed empirical parameters (the PT coefficient, the HS temperature range coefficient and the uncalibrated Thornthwaite adaptation) were not tuned to the local climate. These constraints also dictate the appropriate application of the results. Where complete and reliable meteorological observations are available FAO-56 PM is the preferred standardized estimate [4]. Where radiation, temperature and humidity are available but wind speed is missing PT as implemented here is the most promising alternative for exploratory radiation-based estimates (bearing in mind its local bias). It should not be used for any operational irrigation scheduling without independent local validation. The use of PT without humidity data would require an alternative estimate of net longwave radiation (e.g., actual vapour pressure derived from Tmin [4]) which was not evaluated in this study. Uncalibrated HS may be of some use in a preliminary assessment where only temperature data are available, though its positive bias relative to PM would require local calibration. The simplified BC and daily Thornthwaite are not recommended for any operational daily irrigation scheduling due to their simplicity. Future work should focus on independent lysimeter or flux measurements, station-to-station validation of the RF–ERA5 procedure, uncertainty propagation from the hourly reconstruction to the monthly and annual ETo and local calibration evaluated on a withheld period. This would go some way toward determining whether the degree of agreement documented here translates into better estimates of actual atmospheric water demand and hence better management decisions.

5. Conclusions

We evaluated four simplified ETo methods against the FAO-56 Penman–Monteith method at a humid tropical station with an incomplete record, and tracked the source of each meteorological input. The main findings are as follows:

- **Data provenance.** Of 2713 complete days, only 527 (19.4%) were fully QC-observed, with none in 2021–2024. ETo results for those years are reanalysis-assisted estimates.
- **Reconstruction.** The hybrid RF–ERA5 approach reduced blocked-validation RMSE by 5–15% compared to calendar-only estimates, 10–42% compared to raw ERA5-Seamless, and 7–21% compared to monthly linear bias correction. Comparable performance was observed when entire years were withheld. However, this approach did not preserve Tmin covariation with other variables, and reconstructed inputs produced daily PM errors of about 1 mm/day.
- **Daily agreement.** Priestley–Taylor, using the same net radiation as PM, showed the closest agreement with PM (RMSE = 0.54 mm/day, 95% CI 0.49–0.60; PBIAS = +6.5%; NSE = 0.71; d = 0.92). Hargreaves–Samani and the daily Thornthwaite adaptation both exceeded PM by about 20% (RMSE ≈ 1.0 mm/day) and were statistically indistinguishable. The simplified Blaney–Criddle method exceeded PM by 44% (NSE = −2.49).
- **Monthly agreement.** Priestley–Taylor (RMSE = 14.8 mm/month) and conventional Thornthwaite (14.9 mm/month) were statistically indistinguishable (paired difference 95% CI −4.3 to 3.1 mm/month). Thornthwaite had a lower MAE and higher correlation coefficient. Neither method was superior at the monthly scale.
- **Robustness.** Priestley–Taylor ranked first in all data-provenance scenarios and seasons (RMSE = 0.68 mm/day; PBIAS = −1.0% on fully QC-observed days). The ranking was consistent whether ETo was computed from observed or reconstructed inputs on the same validation days.
- **Formulation and timescale.** The daily Pereira–Pruitt adaptation of Thornthwaite produced monthly totals about 10% higher than the conventional formulation, with its lower effective-temperature bound active on 93.6% of days. This difference is mainly due to using (Tmax + Tmin)/2 instead of the mean of hourly temperatures. The two formulations are not interchangeable near the equator.
- **Practical implications.** Where complete data are available, FAO-56 PM remains the standard. Priestley–Taylor, as implemented here, is a potential exploratory alternative when radiation, temperature, and humidity are available but wind speed is missing. However, it should not be used for operational irrigation scheduling without independent local validation. Uncalibrated Hargreaves–Samani, simplified Blaney–Criddle, and daily Thornthwaite methods require local calibration.
- **Scope.** The findings are based on data from a single station and a model-to-model benchmark. They cannot be generalized to the Colombian Orinoquía or other humid tropical regions without multi-station validation and independent evapotranspiration measurements. ENSO-related differences were exploratory and confounded by data availability.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/hydrology13110313/s1, Table S1: Sensitivity of the out-of-bag (OOB) RMSE of the station-based Random Forest models to the number of trees (maximum depth 15, minimum leaf size 4, max_features 0.8, random_state 42); Table S2: Global radiation (Rs) record before, during and after the quarantine (19 October 2021–16 November 2024): hourly accounting, QC decisions and comparability with ERA5-Seamless; Figure S1: Daily global radiation at La Libertad: raw (unfiltered) station values, QC-accepted station values and ERA5-Seamless, July 2018–December 2025. The shaded band marks the radiation quarantine (19 October 2021–16 November 2024), during which 88.2% of the hourly station values (85.8% of daytime records) were exactly zero; Table S3: Withheld validation blocks and block-level RMSE of the four reconstruction approaches (independent re-run with all steps refitted within folds); Table S4: Robustness of the reconstruction to long outages: leave-one-year-out (LOYO) experiments and 120- and 180-day contiguous blocks. RMSE (MBE in parentheses); Table S5: Pearson correlation matrix of daily inputs within the withheld validation

blocks: observed (QC) values (lower triangle) and reconstructed values (upper triangle); Table S6: ETo computed from observed (QC) and reconstructed inputs on the same withheld validation days (n = 114): (a) error propagated to each method; (b) ranking of the alternative methods relative to PM; Table S7: Daily agreement with FAO-56 PM by data-provenance scenario and season, with 95% moving-block bootstrap intervals of the RMSE; Table S8: Decomposition of the difference between the daily (Pereira–Pruitt) and conventional monthly Thornthwaite formulations (complete series, 89 months); Table S9: Monthly NOAA CPC Relative Oceanic Niño Index (RONI, °C; ERSST version 6, snapshot of 20 October 2026) and ENSO phase assigned to each month of the study period (E: El Niño; L: La Niña; N: neutral); Figure S2: Monthly FAO-56 Penman–Monteith reference evapotranspiration (upper panel) and NOAA RONI (lower panel), with the descriptive classification of El Niño, La Niña and ENSO-neutral months; Figure S3: Distributions of monthly FAO-56 Penman–Monteith anomalies (relative to the 2019–2025 calendar-month mean) grouped by ENSO phase (15 El Niño, 23 neutral and 51 La Niña months). Exploratory and descriptive; months are serially dependent and phase is confounded with data provenance (Table 6 of the main text).

Author Contributions: Conceptualization, C.J.M.C.; methodology, C.J.M.C.; software, C.J.M.C. and A.F.R.V.; validation, C.J.M.C. and A.F.R.V.; formal analysis, A.F.R.V. and A.J.P.Q.; investigation, C.J.M.C. and A.J.P.Q.; resources, A.J.P.Q.; data curation, C.J.M.C.; writing—original draft preparation, C.J.M.C.; writing—review and editing, C.J.M.C., A.J.P.Q. and A.F.R.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The raw hourly station records are available from IDEAM through the DHIME platform (http://dhime.ideam.gov.co, accessed on 16 November 2026), and the ERA5-Seamless predictors are available from the Open-Meteo Historical Weather API [29].

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

References

1. Fietz, C.R.; Silva, F.C.; Urchei, M.A. Estimativa Da Evapotranspiração de Referência Diária Para a Região de Dourados, MS. *Rev. Bras. De Agrometeorol.* **2005**, *13*, 250–255.
2. Bernardo, S.; Soares, A.A.; Mantovani, E.C. *Manual de Irrigação*; Editora UFV: Viçosa, Brazil, 2006.
3. Mendoza Castiblanco, C.J.; Peña Quiñones, A.J. Reference Evapotranspiration Estimation by Different Methods for the Sucroenergy Sector of Colombia. *Rev. Bras. De Eng. Agrícola E Ambient.* **2021**, *25*, 583–590. [CrossRef]
4. Allen, R.G.; Pereira, L.S.; Raes, D.; Smith, M. *Crop Evapotranspiration: Guidelines for Computing Crop Water Requirements*; FAO Irrigation and Drainage Paper; Food and Agriculture Organization of the United Nations: Rome, Italy, 1998; ISBN 92-5-104219-5.
5. Jensen, M.E.; Burman, R.D.; Allen, R.G. *Evapotranspiration and Irrigation Water Requirements*; ASCE Manuals and Reports on Engineering Practice; American Society of Civil Engineers: New York, NY, USA, 1990; ISBN 0-87262-763-2.
6. Borges Júnior, J.C.F.; Anjos, R.J.; Silva, T.J.A.; Lima, J.R.S.; Andrade, C.L.T. Métodos de Estimativa Da Evapotranspiração de Referência Diária Para a Microrregião de Garanhuns, PE. *Rev. Bras. De Eng. Agrícola E Ambient.* **2012**, *16*, 380–390. [CrossRef]
7. Thornthwaite, C.W. An Approach toward a Rational Classification of Climate. *Geogr. Rev.* **1948**, *38*, 55–94. [CrossRef]
8. Conceição, M.A.F.; Marin, F.R. Estimativa Da Evapotranspiração de Referência Utilizando Os Métodos de Hargreaves-Samani e Do Tanque Classe A. In *Proceedings of the Anais do 14º Congresso Brasileiro de Agrometeorologia*; Sociedade Brasileira de Agrometeorologia: Campinas, Brazil, 2005.
9. Mendoza, C.J.; Menezes, R.H.; Dias, A.S. Estimativa Da Evapotranspiração de Referência Por Diferentes Métodos Para o Município de São Luís—MA. *Rev. Bras. De Agric. Irrig.* **2016**, *10*, 621–630. [CrossRef]
10. Rueda Cadavid, J.F.; Cañón Castillo, D.M.; Mendoza Castiblanco, C.J.; Rodríguez Vásquez, A.F.; Cortés Bello, C.A. Evaluating Machine Learning Models for Estimating Evapotranspiration in Colombia's Cauca River Valley. *Sci. Rep.* **2026**, *16*, 278. [CrossRef]
11. Pereira, A.R.; Pruitt, W.O. Adaptation of the Thornthwaite Scheme for Estimating Daily Reference Evapotranspiration. *Agric. Water Manag.* **2004**, *66*, 251–257. [CrossRef]
12. Herrera, W.; Bedoya, O.; Rincón, M. Aplicación de Redes Neuronales Para La Reconstrucción de Series de Tiempo de Precipitación y Temperatura Utilizando Información Satelital. *Rev. EIA* **2020**, *17*, 1–16. [CrossRef]
13. Jimenez, V.; Will, A.; Rodriguez, S.; Lamelas, C. Imputación de Datos Climáticos Utilizando Algoritmos Genéticos Niching. In Proceedings of the XXXVII Reunión de Trabajo de la Asociación Argentina de Energías Renovables y Medio Ambiente (ASADES), Oberá, Argentina, 28–31 October 2014; Volume 2, pp. 11139–11148.

14. Kim, S.-J.; Bae, S.-J.; Jang, M.-W. Linear Regression Machine Learning Algorithms for Estimating Reference Evapotranspiration Using Limited Climate Data. *Sustainability* **2022**, *14*, 11674. [CrossRef]
15. Triana-Madrid, J.C.; Ocampo-Marulanda, C.; Carvajal-Escobar, Y.; Torres-López, W.A.; Triana, J.; Canchala, T. Estimation of Monthly Reference Evapotranspiration with Scarce Information Using Machine Learning in Southwestern Colombia. *Meteorologica* **2023**, *48*, 024. [CrossRef]
16. Das, S.; Baweja, S.K.; Raheja, A.; Gill, K.K.; Sharda, R. Development of Machine Learning-Based Reference Evapotranspiration Model for the Semi-Arid Region of Punjab, India. *J. Agric. Food Res.* **2023**, *13*, 100640. [CrossRef]
17. Považanová, B.; Čistý, M.; Bajtek, Z. Using Feature Engineering and Machine Learning in FAO Reference Evapotranspiration Estimation. *J. Hydrol. Hydromech.* **2023**, *71*, 425–438. [CrossRef]
18. Aly, M.S.; Darwish, S.M.; Aly, A.A. High Performance Machine Learning Approach for Reference Evapotranspiration Estimation. *Stoch. Environ. Res. Risk Assess.* **2024**, *38*, 689–713. [CrossRef]
19. Chen, Z.; Zhu, Z.; Jiang, H.; Sun, S. Estimating Daily Reference Evapotranspiration Based on Limited Meteorological Data Using Deep Learning and Classical Machine Learning Methods. *J. Hydrol.* **2020**, *591*, 125286. [CrossRef]
20. Bellido-Jiménez, J.A.; Estévez, J.; García-Marín, A.P. New Machine Learning Approaches to Improve Reference Evapotranspiration Estimates Using Intra-Daily Temperature-Based Variables in a Semi-Arid Region of Spain. *Agric. Water Manag.* **2021**, *245*, 106558. [CrossRef]
21. Hersbach, H.; Bell, B.; Berrisford, P.; Hirahara, S.; Horányi, A.; Muñoz-Sabater, J.; Nicolas, J.; Peubey, C.; Radu, R.; Schepers, D.; et al. The ERA5 Global Reanalysis. *Q. J. R. Meteorol. Soc.* **2020**, *146*, 1999–2049. [CrossRef]
22. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
23. Stekhoven, D.J.; Bühlmann, P. MissForest—Non-Parametric Missing Value Imputation for Mixed-Type Data. *Bioinformatics* **2012**, *28*, 112–118. [CrossRef]
24. De Los Ríos, K.; Torres-Castillo, J.R.; Rincón-Mejía, W.J.; Celis-Montealegre, E.R.; Riaño-Rivera, A.J.; Gómez-Heredia, C.L. Regional Hydroclimatic Sensitivity of Monthly Precipitation Anomalies to ENSO in the Colombian Andes and Orinoquia. *Hydrology* **2026**, *13*, 223. [CrossRef]
25. Instituto de Hidrología; Meteorología y Estudios Ambientales (IDEAM). *Promedios Climatológicos 1981–2010*; IDEAM: Bogotá, Colombia, 2014.
26. NASA/METI/AIST/Japan Spacesystems; U.S./Japan ASTER Science Team. *ASTER Global Digital Elevation Model V003*; NASA EOSDIS Land Processes Distributed Active Archive Center: Sioux Falls, SD, USA, 2019. [CrossRef]
27. Farr, T.G.; Rosen, P.A.; Caro, E.; Crippen, R.; Duren, R.; Hensley, S.; Kobrick, M.; Paller, M.; Rodriguez, E.; Roth, L.; et al. The Shuttle Radar Topography Mission. *Rev. Geophys.* **2007**, *45*, RG2004. [CrossRef]
28. Probst, P.; Boulesteix, A.-L. To Tune or Not to Tune the Number of Trees in Random Forest. *J. Mach. Learn. Res.* **2018**, *18*, 1–18.
29. Zippenfenig, P. *Open-Meteo.com Weather API*, Version 1.4.0; Zenodo: Geneva, Switzerland, 2023. [CrossRef]
30. 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]
31. Doorenbos, J.; Pruitt, W.O. *Guidelines for Predicting Crop Water Requirements*; FAO Irrigation and Drainage Paper 24; Food and Agriculture Organization of the United Nations: Rome, Italy, 1977.
32. Willmott, C.J.; Rowe, C.M.; Mintz, Y. Climatology of the Terrestrial Seasonal Water Cycle. *J. Climatol.* **1985**, *5*, 589–606. [CrossRef]
33. de Camargo, A.P.; Sentelhas, P.C. Avaliação Do Desempenho de Diferentes Métodos de Estimativa Da Evapotranspiração Potencial No Estado de São Paulo, Brasil. *Rev. Bras. De Agrometeorol.* **1997**, *5*, 89–97.
34. Willmott, C.J.; Ackleson, S.G.; Davis, R.E.; Feddema, J.J.; Klink, K.M.; Legates, D.R.; O'Donnell, J.; Rowe, C.M. Statistics for the Evaluation and Comparison of Models. *J. Geophys. Res.* **1985**, *90*, 8995–9005. [CrossRef]
35. Nash, J.E.; Sutcliffe, J.V. River Flow Forecasting through Conceptual Models Part I—A Discussion of Principles. *J. Hydrol.* **1970**, *10*, 282–290. [CrossRef]
36. Künsch, H.R. The Jackknife and the Bootstrap for General Stationary Observations. *Ann. Stat.* **1989**, *17*, 1217–1241. [CrossRef]
37. Diebold, F.X.; Mariano, R.S. Comparing Predictive Accuracy. *J. Bus. Econ. Stat.* **1995**, *13*, 253–263. [CrossRef]
38. NOAA Climate Prediction Center Relative Oceanic Niño Index (RONI): Historical El Niño/La Niña Episodes. Available online: https://www.cpc.ncep.noaa.gov/products/analysis_monitoring/enso/roni/ (accessed on 20 October 2026).
39. Pereira, A.R. The Priestley–Taylor Parameter and the Decoupling Factor for Estimating Reference Evapotranspiration. *Agric. For. Meteorol.* **2004**, *125*, 305–313. [CrossRef]
40. Trajkovic, S. Hargreaves versus Penman–Monteith under Humid Conditions. *J. Irrig. Drain. Eng.* **2007**, *133*, 38–42. [CrossRef]
41. Droogers, P.; Allen, R.G. Estimating Reference Evapotranspiration under Inaccurate Data Conditions. *Irrig. Drain. Syst.* **2002**, *16*, 33–45. [CrossRef]
42. Pimentel, R.; Arheimer, B.; Crochemore, L.; Andersson, J.C.M.; Pechlivanidis, I.G.; Gustafsson, D. Which Potential Evapotranspiration Formula to Use in Hydrological Modeling World-Wide? *Water Resour. Res.* **2023**, *59*, e2022WR033447. [CrossRef]

43. Poveda, G.; Jaramillo, A.; Gil, M.M.; Quiceno, N.; Mantilla, R.I. Seasonality in ENSO-Related Precipitation, River Discharges, Soil Moisture, and Vegetation Index in Colombia. *Water Resour. Res.* **2001**, *37*, 2169–2178. [CrossRef]
44. Poveda, G.; Álvarez, D.M.; Rueda, Ó.A. Hydro-Climatic Variability over the Andes of Colombia Associated with ENSO: A Review of Climatic Processes and Their Impact on One of the Earth's Most Important Biodiversity Hotspots. *Clim. Dyn.* **2011**, *36*, 2233–2249. [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.