

Downscaling dekadal WaPOR actual evapotranspiration from 300 m to 20 m for irrigation monitoring

Zolo Kiala

Z.Kiala@cgiar.org

International Water Management Institute

Schmitter Petra

International Water Management Institute

Mariangel Garcia Andarcia

International Water Management Institute

Research Article

Keywords: evapotranspiration, downscaling, WaPOR, CatBoost, Sentinel-2, irrigation monitoring

Posted Date: June 8th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9687804/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Downscaling dekadal WaPOR actual evapotranspiration from 300 m to 20 m for irrigation monitoring

Zolo Kiala^{1*}, Schmitter Petra², Mariangel Garcia Andarcia²,

¹*International Water Management Institute, Block G Ground Floor, Hatfield Gardens, 333 Grosvenor St, Hatfield, Pretoria, 0083, South Africa*

²*International Water Management Institute, 127 Sunil Mawatha, Battaramulla, Sri Lanka 10120*

*Correspondence: Z.Kiala@cgiar.org

Abstract

Irrigated agriculture accounts for nearly 70% of global freshwater withdrawals, yet monitoring consumptive water use at field scale remains challenging in fragmented smallholder systems. FAO's WaPOR database provides global dekadal actual evapotranspiration (ETa) at approximately 300 m resolution, but this scale cannot resolve sub-field heterogeneity typical of southern Africa. This study presents WaPOR-AETI-DownScaler v1.0, an operational framework for downscaling WaPOR ETa from 300 m to 20 m while assessing whether large-scale magnitude consistency is maintained. The approach applies CatBoost regression using Sentinel-2 reflectance and vegetation indices (NDVI, NDMI, FVC), harmonic seasonal encoding, CHIRPS rainfall, terrain variables, and coarse-resolution WaPOR ETa as a structural predictor. Model development over two WaPOR Level-3 pilot areas in Mozambique used approximately 500,000 dekadal pixel samples and 150 feature-subset optimization trials. Internal validation, used for model selection, achieved $R^2 = 0.883$, $RMSE = 3.95 \text{ mm dekad}^{-1}$, and $RRMSE = 19.5\%$. Independent temporal testing on hold-out dekads, used for generalization assessment, maintained stable pixel-level performance (mean $pvn_r^2 \approx 0.62$; relative $RMSE \approx 23\text{--}24\%$). No statistically significant difference was detected between aggregated 20 m predictions and WaPOR Level-1 ETa at 300 m ($p > 0.05$), indicating no detectable degradation of aggregated 300 m ETa magnitude after downscaling. The framework enhances fine-scale ETa heterogeneity without detectable degradation of basin-scale magnitude, providing a scalable tool for irrigation monitoring and water accounting in data-scarce regions.

Keywords: evapotranspiration · downscaling · WaPOR · CatBoost · Sentinel-2 · irrigation monitoring

1 Introduction

Irrigated agriculture is the largest consumptive user of freshwater globally, accounting for about 70% of total withdrawals (FAO 2017; WWAP 2021). Reliable quantification of agricultural water consumption is therefore critical for basin-scale water accounting, drought monitoring, irrigation performance assessment, and sustainable water-resources management (Molden et al. 2010; Karimi et al. 2013). In sub-Saharan Africa, irrigation systems are often fragmented, heterogeneous, and poorly documented, limiting conventional inventory-based approaches (You et al. 2011; Xie et al. 2014). Satellite remote sensing has consequently become a central tool for monitoring agricultural water use at regional scales (Bastiaanssen et al. 1998; Anderson et al. 2012; Senay et al. 2013).

Satellite-derived evapotranspiration (ET) provides a physically meaningful measure of consumptive water use. Field-scale energy-balance models using Landsat thermal data (e.g., SEBAL, METRIC) demonstrate strong accuracy but are constrained by revisit frequency and cloud cover (Bastiaanssen et al. 1998; Allen et al. 2007). Coarser-resolution products such as MOD16 and GLEAM offer temporal consistency at regional and global scales but lack spatial detail (Mu et al. 2011; Martens et al. 2017). Multi-scale systems such as ALEXI/DisALEXI and ECOSTRESS further improve spatial variability mapping, yet operational coverage remains limited (Anderson et al. 2018; Fisher et al. 2020).

The FAO WaPOR database represents a major advancement for operational water accounting in data-scarce regions, providing harmonized actual evapotranspiration (ETa, equivalent to WaPOR AETI) estimates at dekadal and annual time steps (FAO 2018; Blatchford et al. 2020). However, WaPOR Level-1 and Level-2 AETI are available at ~250–300 m resolution, which is insufficient to resolve fragmented smallholder systems and heterogeneous field mosaics common in southern Africa. Although Level-3 provides 20 m AETI, it is limited to selected pilot areas.

To address this spatial gap, recent studies have applied statistical and machine-learning approaches to downscale coarse ETa products using high-resolution predictors. Ensemble tree-based models such as Random Forest and Gradient Boosting have shown strong performance in capturing non-linear relationships between ETa and multispectral, terrain, and climatic variables (Xu et al. 2018). Deep learning approaches have also been explored but often require large training datasets and substantial computational resources, limiting operational scalability (Sun et al. 2019; Che et al. 2022). Moreover, many studies focus on seasonal or annual

aggregates, while relatively few address multi-year downscaling of dekadal ETa, which is particularly relevant for near-real-time irrigation monitoring.

In response, we introduce WaPOR-AETI-DownScaler v1.0, an operational framework for downscaling dekadal WaPOR AETI from 300 m to 20 m resolution using CatBoost, a gradient-boosted decision-tree algorithm known for stability and generalization performance. The framework integrates a minimal and robust predictor set derived from Sentinel-2 imagery, terrain variables, seasonal indicators, rainfall, and coarse-resolution AETI. It is developed and evaluated over two WaPOR Level-3 pilot areas in Mozambique, enabling consistent training and validation against 20 m AETI reference data. This study contributes a reproducible and operational framework for downscaling dekadal WaPOR ETa from 300 m to 20 m in fragmented irrigation landscapes. Its novelty lies not only in the use of machine learning for spatial refinement, but in the integration of multi-year dekadal training, independent temporal hold-out evaluation, and explicit testing of whether aggregated 20 m predictions remain consistent with the original WaPOR 300 m ETa signal. This combination supports both field-scale irrigation diagnostics and basin-scale water accounting.

2 Materials and Methods

2.1 Study area description

The study was conducted in two irrigated agricultural regions in southern Mozambique: the Baixo Limpopo Irrigation Scheme and the Lamego Irrigated Area, both located within the Limpopo River Basin (Figure 1). These sites were selected because they are covered by FAO



Figure 1 Study areas used for the development and evaluation of WaPOR-AETI-DownScaler v1.0.

WaPOR Level-3 evapotranspiration (ETa) products at 20 m spatial resolution, which provide high-resolution reference data suitable for model development and validation (FAO 2018; Blatchford et al. 2020).

2.1.1 Baixo Limpopo Irrigation Scheme

The Baixo Limpopo Irrigation Scheme is located in Gaza Province ($\approx 25.0^\circ$ S, 33.6° E). The irrigated area ranges between 35,000–45,000 ha (≈ 350 – 450 km²), depending on season and operational infrastructure. The climate is semi-arid to sub-humid, with annual rainfall between 500–800 mm concentrated during November–March. Potential evapotranspiration exceeds rainfall during much of the year, making irrigation essential. Cropping systems include rice, maize, sugarcane, and horticultural crops. The terrain is predominantly flat, facilitating surface irrigation but increasing sensitivity to water management.

2.1.2 Lamego Irrigated Area

The Lamego irrigated area ($\approx 24.3^\circ$ S, 33.9° E) covers approximately 3,000–6,000 ha (≈ 30 – 60 km²). Agriculture is dominated by fragmented smallholder systems. Fields are heterogeneous and irrigation intensity varies spatially. Terrain and soil variability contribute to stronger spatial ET heterogeneity compared to Baixo, making Lamego a suitable contrast site for robustness assessment.

2.2 Downscaling Framework

WaPOR-AETI-DownScaler v1.0 downscales WaPOR Level-1 ETa (~ 300 m) to 20 m resolution using a machine-learning approach. The workflow consists of:

- Multi-source data preprocessing in Google Earth Engine
- Feature engineering at dekadal resolution
- Pixel sampling from WaPOR Level-3 ET stacks
- CatBoost model training and feature subset selection
- Independent temporal hold-out evaluation
- Multi-scale aggregation and consistency analysis

All predictors (as shown in table 1) were spatially aligned to the WaPOR Level-3 20 m grid.

2.2.1 Data Preparation

All datasets were aggregated to dekadal (10-day) resolution.

Sentinel-2

Sentinel-2 surface reflectance imagery (Drusch et al. 2012) was filtered using a maximum cloud threshold of 40%. Cloud and shadow masking was performed using the Scene Classification Layer (Louis et al. 2016). A median composite was generated per dekad to reduce residual contamination.

WaPOR Evapotranspiration

WaPOR Level-1 ET (~300 m) was aggregated per dekad and used as a coarse-resolution predictor (FAO 2018). WaPOR Level-3 ETa (20 m) was extracted for the same dekads and used as the training target (Blatchford et al. 2020). Both products represent actual evapotranspiration in millimetres per dekad.

Rainfall and Meteorology

Climate Hazards Center InfraRed Precipitation with Station data (CHIRPS) precipitation (Funk et al. 2015) and ECMWF Reanalysis v5 (ERA5)-Land meteorological variables (Muñoz-Sabater et al. 2021) were aggregated to dekadal values.

Terrain and Land Cover

Elevation data were sourced from the Copernicus DEM (Copernicus DEM 2020), while land-cover information was obtained from the European Space Agency (ESA) WorldCover dataset (Zanaga et al. 2022). Continuous predictors were resampled using bilinear interpolation, whereas categorical variables were resampled using nearest-neighbour mode. Dekads with missing Sentinel-2 composites, Level-1 ETa predictors, or Level-3 ETa targets were excluded from the analysis.

2.2.2 Feature Engineering

Feature engineering transformed raw satellite and reanalysis inputs into predictors relevant to evapotranspiration processes.

Optical and Vegetation Indices

Spectral bands B2–B12 were retained. Vegetation and moisture indices included:

- Normalized Difference Vegetation Index (NDVI) (Rouse et al. 1974)
- Normalized Difference Moisture Index (NDMI) (Gao 1996)
- Normalized Burn Ratio (NBR) (Key & Benson 2006)

- Enhanced Vegetation Index (EVI) (Huete et al. 2002)
- Soil-Adjusted Vegetation Index (SAVI) (Huete 1988)

Fractional vegetation cover (FVC) was computed from NDVI (Carlson & Ripley 1997).

Rainfall and Meteorology

Rainfall predictors included dekadal totals, lagged totals, and rolling 30- and 60-day accumulations (Funk et al. 2015). Meteorological predictors included air temperature, vapour pressure deficit, shortwave radiation, wind speed, and relative humidity (Muñoz-Sabater et al. 2021).

Terrain and Surface Context

Elevation, slope, and aspect were derived from the Digital Elevation Model (DEM). Aspect was transformed into sine and cosine components to avoid angular discontinuities (Wilson & Gallant 2000).

Seasonal Encoding

Seasonality was represented using harmonic encoding:

$$SIN_DOY = \sin\left(\frac{2\pi \times DOY}{365}\right) \quad \text{Eq. (1)}$$

$$COS_DOY = \cos\left(\frac{2\pi \times DOY}{365}\right) \quad \text{Eq. (2)}$$

This captures annual periodicity without introducing temporal discontinuities (Beer et al. 2010). Tree-based models do not require feature normalization.

Table 1 Summary of input features and training target used for evapotranspiration downscaling

Category	Dataset	Variables	Native resolution	Temporal scale	Reference
Optical	Sentinel-2 SR	B2–B12, NDVI, NDMI, NBR, EVI, SAVI, FVC	10–20 m	Dekadal	Drusch et al. (2012)
ET (predictor)	FAO WaPOR Level-1	ETa300m	250–300 m	Dekadal	FAO (2018)
ET (target)	FAO WaPOR Level-3	ETa20m	20 m	Dekadal	Blatchford et al. (2020)

Category	Dataset	Variables	Native resolution	Temporal scale	Reference
Rainfall	CHIRPS	10d, lag-10d, 30d, 60d sums	~5 km	Dekadal	Funk et al. (2015)
Meteorology	ERA5-Land	T2M, VPD, SWDOWN, WIND, RH	~9 km	Dekadal	Muñoz-Sabater et al. (2021)
Terrain	Copernicus DEM	Elevation, slope, aspect (sin, cos)	30 m	Static	Copernicus DEM (2020)
Land cover	ESA WorldCover	Land-cover class	10 m	Static	Zanaga et al. (2022)
Seasonality	DOY encoding	SIN_DOY, COS_DOY	—	Dekadal	Beer et al. (2010)

2.2.3 Model Development and Optimization

Training Data Construction

Training samples were extracted from temporally stacked WaPOR Level-3 predictor layers using site-specific temporal windows (Baixo: 1 January 2018–16 December 2022; Lamego: 21 January 2019–20 December 2020) to capture interannual variability in evapotranspiration dynamics. All GeoTIFF stacks were screened to ensure the presence of required predictor and target bands prior to sampling. Approximately 500,000 samples were generated using a controlled spatial sampling strategy, with balanced representation across files (minimum 500 and maximum 4,000 samples per file). To improve robustness, pixels with slope $> 25^\circ$ were excluded, and sampling was biased toward vegetated areas (70% of samples) defined by NDVI > 0.35 or FVC > 0.20 . Additional filtering removed missing or invalid values, negative ETa values, and non-physical extremes. Each sample consisted of a feature vector $\mathbf{x}_i \in \mathbb{R}^p$, derived from multispectral, climatic, seasonal, rainfall, coarse-resolution ETa, and terrain predictors, and a corresponding target value ET_i^{20m} from WaPOR Level-3 ETa.

Data Partitioning

The dataset was randomly partitioned into 85% training and 15% validation subsets. The validation set was used exclusively for feature-group subset selection and hyperparameter tuning to prevent information leakage and ensure unbiased model optimisation. Model optimisation was performed using 150 feature-subset trials, with a minimum of two predictor groups retained per configuration.

Because training and validation samples were drawn from spatially structured pixels, residual spatial autocorrelation between neighbouring samples may inflate internal validation metrics

relative to fully independent locations. To mitigate this, the final assessment of generalization performance was based on per-file temporal hold-out dekads outside the site-specific model-development windows, which are temporally independent from the training data and therefore less affected by within-site spatial autocorrelation.

Learning Algorithm

Evapotranspiration was modelled using CatBoost, a gradient boosting framework based on symmetric (oblivious) decision trees (Prokhorenkova et al. 2018). CatBoost minimizes a differentiable loss function using ordered boosting to reduce target leakage and prediction shift. CatBoost was selected over alternative approaches such as Random Forest, XGBoost, and deep-learning regressors because its built-in regularization, native handling of categorical features, and robustness to feature scaling deliver stable performance on high-dimensional, spatially heterogeneous geospatial inputs at substantially lower computational cost than neural-network alternatives.

For regression, the objective function is:

$$\mathcal{L} = \frac{1}{N} \sum_{i=1}^N \ell \left(ET_i^{20m}, \widehat{ET}_i^{20m} \right) \quad \text{Eq. (3)}$$

where ℓ denotes the loss function (e.g., RMSE loss), and predictions are generated as:

$$\widehat{ET}_i^{20m} = \sum_{t=1}^T f_t(\mathbf{x}_i) \quad \text{Eq. (4)}$$

with T denoting the number of boosting iterations and f_t representing individual symmetric decision trees.

Hyperparameter Optimization

Hyperparameters were optimized using the internal validation set. The tuning process explored:

- Tree depth
- Learning rate
- Number of boosting iterations
- L2 leaf regularization
- Subsampling rate

Model selection was based on minimizing validation RMSE while monitoring variance inflation and overfitting behavior.

Final Model Training

After identifying the optimal hyperparameter configuration using the internal validation subset, the final CatBoost model was retrained using the full model-development dataset, comprising both the training and validation samples. This step was performed to maximize the use of available information for parameter estimation before applying the model to independent spatial–temporal evaluation datasets.

The validation subset was used only for feature-group selection and hyperparameter tuning; it was not used to report independent test performance. Generalization performance was assessed separately using per-file hold-out dekads outside the site-specific model-development windows. Specifically, evaluation used dates outside the Baixo training window (1 January 2018–16 December 2022) and outside the Lamego training window (21 January 2019–20 December 2020). This procedure ensured that model selection was conducted without information leakage, while the final reported hold-out evaluation remained temporally independent from the model-development data.

Final Model Configuration

The final model configuration retained all six predictor groups and 15 predictor bands, trained with the optimal CatBoost hyperparameters identified during the subset-search and hyperparameter-tuning stage. These settings, summarised in Table 2, are reported here to support reproducibility.

Table 2 Final WaPOR-AETI-DownScaler v1.0 model configuration, including the 15 selected predictor bands and the CatBoost hyperparameters used to train the production model.

Item	Final value
Predictor groups (6)	doy, eta300, idx, rain, s2, terrain
Predictor bands (15)	ETa300m; NDVI, NDMI, FVC; SIN_DOY, COS_DOY; RAIN_10d, RAIN_10d_lag; B4, B8, B11; DEM, Slope, Aspect_sin, Aspect_cos
Loss function	RMSE
Evaluation metric	RMSE
Iterations	1200

Depth	4
Learning rate	0.1
L2 leaf regularization	1.0
Bootstrap type	Bayesian
Bagging temperature	0.5
Overfitting detector	Iter (od_wait = 40)
Random seed	22
Task type	CPU
CatBoost version	1.2.8

2.2.4 Evaluation Framework

Pixel-Level Evaluation (20 m)

Let:

- ET_i^{20m} denote the native WaPOR Level-3 evapotranspiration at pixel i , expressed in millimetres per dekad, and
- $\hat{ET}_i^{20m}$ denote the corresponding predicted evapotranspiration generated by the downscaling model at 20 m resolution.

Model performance was quantified using five complementary statistical metrics: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Bias, Relative RMSE (RRMSE), and the coefficient of determination (R^2).

The Root Mean Square Error (RMSE) measures the overall magnitude of prediction errors and gives greater weight to larger deviations. It is defined as:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N \left(\hat{ET}_i^{20m} - ET_i^{20m} \right)^2} \quad \text{Eq. (5)}$$

Where:

- N is the number of valid pixels,
- ET_i^{20m} is the native WaPOR Level-3 evapotranspiration at pixel i , and

- $\hat{ET}_i^{20m}$ is the predicted evapotranspiration at the same pixel i , and ET_i^{20m} is the predicted evapotranspiration at the same pixel.

The Mean Absolute Error (MAE) provides a robust measure of the average absolute deviation between predicted and observed values:

$$MAE = \frac{1}{N} \sum_{i=1}^N \left| \hat{ET}_i^{20m} - ET_i^{20m} \right| \quad \text{Eq. (6)}$$

The Bias quantifies systematic overestimation or underestimation:

$$Bias = \frac{1}{N} \sum_{i=1}^N \left(\hat{ET}_i^{20m} - ET_i^{20m} \right) \quad \text{Eq. (7)}$$

A positive bias indicates overestimation, while a negative bias indicates underestimation.

The Relative RMSE (RRMSE) expresses the RMSE as a percentage of the mean observed evapotranspiration:

$$RRMSE = \frac{RMSE}{\overline{ET}^{20m}} \times 100 \quad \text{Eq. (8)}$$

where $\overline{ET}^{20m}$ is the mean of the observed WaPOR Level-3 ETa values. This normalization facilitates comparison across sites and seasons with different evapotranspiration magnitudes.

Finally, the coefficient of determination (R^2) measures the proportion of variance in the observed ETa explained by the model:

$$R^2 = 1 - \frac{\sum_{i=1}^N \left(\hat{ET}_i^{20m} - ET_i^{20m} \right)^2}{\sum_{i=1}^N \left(ET_i^{20m} - \overline{ET}^{20m} \right)^2} \quad \text{Eq. (9)}$$

Together, these metrics provide a comprehensive assessment of agreement with WaPOR Level-3 ETa, systematic bias, and variance explanation at the 20 m spatial resolution.

Multi-Scale Aggregation (20 m $\rightarrow$ 300 m)

Evaluation was conducted per dekad and per site to assess both temporal and spatial model performance.

For large-scale consistency analysis, the native 20 m evapotranspiration estimates were spatially aggregated to the WaPOR 300 m grid:

$$ET_{300m}^{agg} = \frac{1}{M} \sum_{j=1}^M \widehat{ET}_j^{20m} \quad \text{Eq. (10)}$$

where M represents the number of 20 m pixels within each corresponding 300 m WaPOR pixel. These aggregated values were then compared against the original WaPOR Level-1 evapotranspiration product at 300 m resolution:

$$ET_{300m}^{WaPOR} \quad \text{Eq. (11)}$$

This aggregation–comparison framework ensures that spatial refinement through downscaling assesses whether large-scale magnitude consistency is maintained, in line with established multi-scale validation approaches (Peng et al. 2017; Yang et al. 2020).

2.2.5 Statistical Uncertainty and Hypothesis Testing

Uncertainty in dekadal performance metrics was quantified using non-parametric bootstrap resampling across dekads. Differences between ETa datasets were then assessed using paired statistical tests to evaluate whether aggregated Pred20 and Native20 ETa differed significantly from WaPOR Level-1 ETa at 300 m.

Bootstrap Confidence Intervals

Uncertainty in mean dekadal metrics (e.g., pvn_r^2 , RMSE, RRMSE) was estimated using bootstrap resampling across dekads. For a given metric m computed over K dekads, bootstrap samples were generated by resampling the dekadal values with replacement:

$$\bar{m}^{(b)} = \frac{1}{K} \sum_{k=1}^K m_k^* \quad \text{Eq. (12)}$$

where m_k^* denotes a resampled dekadal value and $b = 1, \dots, B$ bootstrap iterations ($B = 5000$).

The 95% confidence interval was defined as the 2.5th and 97.5th percentiles of the bootstrap distribution. This non-parametric approach avoids normality assumptions and is robust to skewed error distributions.

Seasonal Comparison

To assess seasonal differences (wet: October–April; dry: May–September), we applied:

- Welch’s two-sample t-test (unequal variances)
- Mann–Whitney U test (non-parametric)

This dual approach ensures robustness to non-normal distributions and unequal variances.

Multi-Scale Conservation Test

To evaluate large-scale consistency after aggregation (20 m $\rightarrow$ 300 m), paired comparisons between Pred20 and Native20 aggregated RMSE were performed using:

- Paired t-test
- Wilcoxon signed-rank test

Lack of significant difference was interpreted as evidence that no detectable degradation occurred after aggregation, rather than as formal statistical equivalence.

2.3 Reproducibility and Deployment

Preprocessing was conducted in Google Earth Engine, and model training was performed in Python using CatBoost. Datasets and models were versioned using Weights & Biases (Biewald 2020). The full implementation of the workflow is publicly available at: <https://github.com/ZoloKiala/WaPOR-downscaling>, ensuring reproducibility. The trained model is deployed as an operational web application at: https://huggingface.co/spaces/IWMIHQ/WaPOR_20_m_Downscaler.

3 Results

3.1 Feature Subset Selection and Internal Performance

3.1.1 Model Development and Internal Validation

A total of 500,000 stratified pixel samples from Baixo (2018–2022) and Lamego (2019–2020) were split into 85% training and 15% validation sets. Model development combined feature-group subset selection and randomized hyperparameter optimization across 150 CatBoost trials.

The optimal configuration retained all six predictor groups (doy, eta300, idx, rain, s2, terrain), yielding 15 predictor bands. Internal validation showed strong performance (RMSE = 3.95 mm dekad⁻¹, MAE = 2.54 mm dekad⁻¹, $R^2 = 0.883$, RRMSE = 19.5%), explaining approximately 88% of pixel-scale ETa variability.

Across all trials, performance varied substantially (RMSE: 3.95–10.34 mm dekad⁻¹; R^2 : 0.20–0.88), with the optimal model clearly outperforming the median configuration (RMSE = 4.82 mm dekad⁻¹; $R^2 = 0.826$). Spatial comparisons for representative high-performing dekads over Lamego and Baixo (Figures 2–3) illustrate the strong agreement between downscaled 20 m ETa and native WaPOR ETa, as well as the improved representation of sub-field variability relative to the aggregated 300 m resolution.

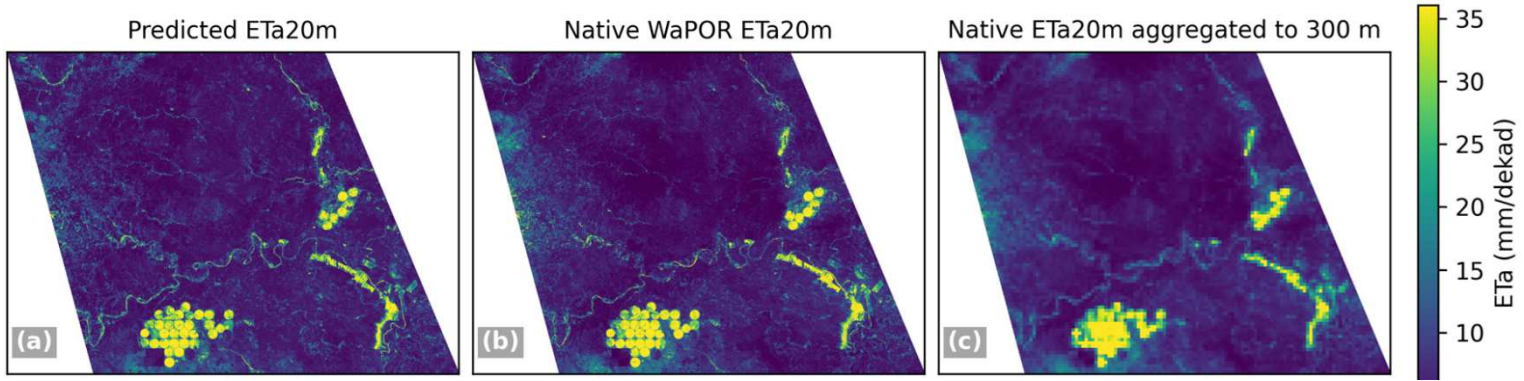


Figure 2 Comparison of evapotranspiration (ETa) over Lamego on 11 November 2020 (maximum pixel-level $R^2 = 0.89$). (a) Downscaled ETa at 20 m resolution, (b) native WaPOR ETa at 20 m, and (c) native ETa aggregated to 300 m and resampled to 20 m.

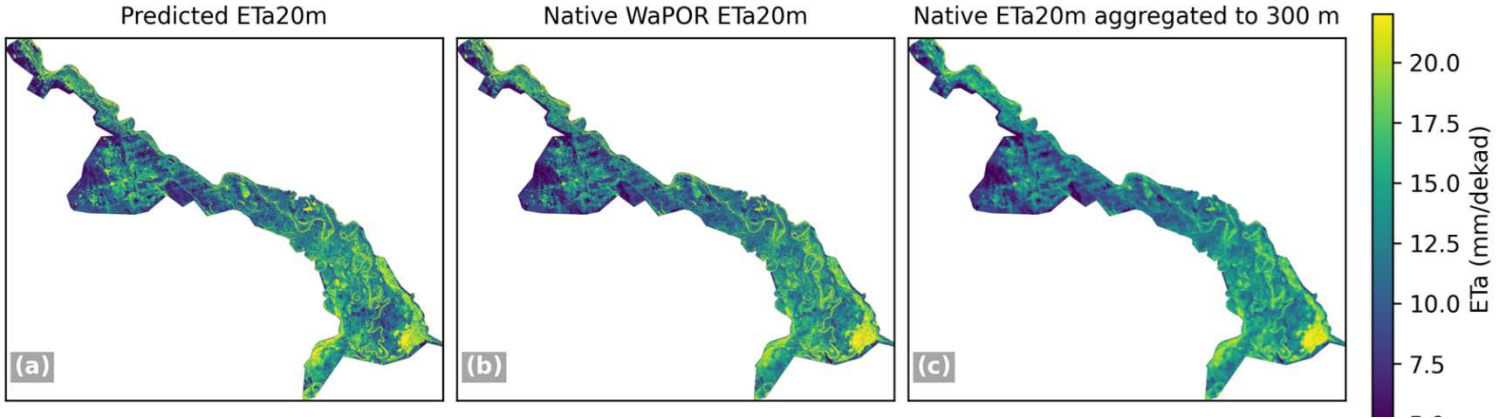


Figure 3 Comparison of evapotranspiration (ETa) over Baixo Limpopo on 4 July 2023 (maximum pixel-level $R^2 = 0.82$). (a) Downscaled ETa at 20 m resolution, (b) native WaPOR ETa at 20 m, and (c) native ETa aggregated to 300 m and resampled to 20 m.

3.1.2 Predictor Importance and Model Drivers

Feature-importance analysis of the selected CatBoost model (Figure 4) indicates that evapotranspiration variability is primarily driven by vegetation moisture status and seasonal dynamics. The highest-ranked predictors were NDMI (13.29), COS_DOY (13.27), NDVI (11.56), SIN_DOY (11.51), ETa300m (10.97), FVC (9.56), RAIN_10d (7.20), and DEM (6.84).

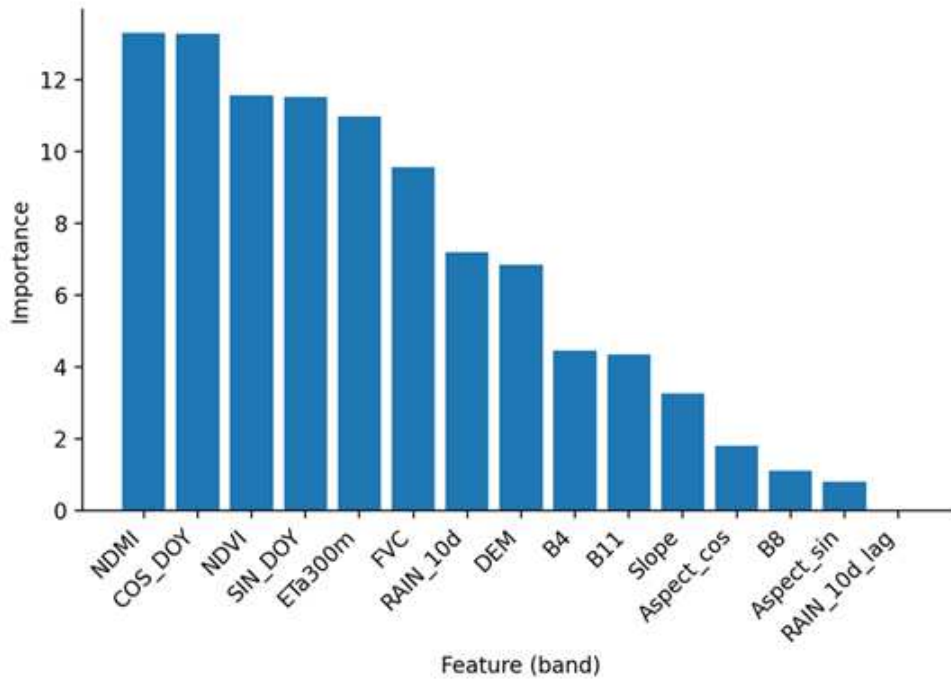


Figure 4 Top 15 feature importances of the selected CatBoost model. Vegetation indices (NDMI, NDVI, FVC) and seasonal harmonics (SIN_DOY, COS_DOY) are the dominant predictors, followed by ETa300m and rainfall, while terrain and individual spectral bands contribute less.

Vegetation indices (NDMI, NDVI, FVC) collectively dominate the ranking, underscoring the importance of canopy moisture and greenness in explaining spatial ET heterogeneity. The harmonic seasonal terms (SIN_DOY, COS_DOY) are similarly influential, confirming that phenological phase strongly regulates evapotranspiration patterns. The prominence of ETa300m highlights the role of coarse-resolution WaPOR ETa as a structural large-scale constraint in the downscaling process.

In contrast, individual reflectance bands (B4, B11, B8) and terrain-derived variables (Slope, Aspect components) contribute comparatively lower importance, suggesting that biophysical condition and seasonal forcing are stronger determinants than raw spectral reflectance alone. Consistent selection of all six feature groups across subset-search trials (Figure 5) further demonstrates that ETa dynamics in these fragmented irrigation systems are inherently multi-factorial and cannot be adequately represented by a reduced predictor set without performance loss.

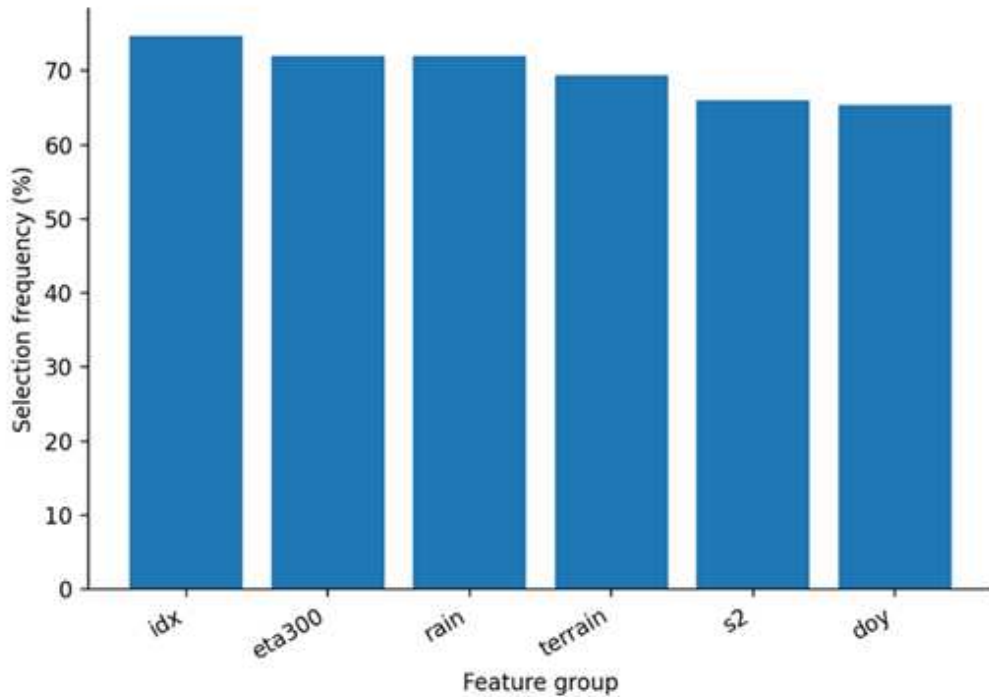


Figure 5 Feature-group selection frequency across subset-search trials. Vegetation indices, ETa300m, and rainfall were selected most frequently, followed by terrain, Sentinel-2 reflectance, and seasonal encoding, highlighting the multi-factorial control.

3.2 Pixel-Level Evaluation (20 m)

Pixel-level evaluation examined the agreement between predicted 20 m evapotranspiration (Pred20) and native WaPOR Level-3 ETa (Native20) using the coefficient of determination (pvn_r^2) and dekad-level error metrics (Figures 6 and 7). Figures 6a–c (Baixo) and 7a–c (Lamego) illustrate the temporal evolution of pvn_r^2 , RMSE/MAE, and relative RMSE (%), respectively. To represent typical conditions, dekads with extremely low explanatory power ($pvn_r^2 < 0.1$) were excluded from bootstrap summaries (2 removed in Baixo; 7 in Lamego).

In Baixo (2023–2024; Figure 6), the temporal series shows generally stable and positive pvn_r^2 values, mostly above 0.6, with one pronounced negative outlier during late August 2023 associated with very low ETa variance. After filtering, the mean pvn_r^2 was 0.62 (95% CI: 0.56–0.68; $n = 31$), with mean pvn_rmse of 4.43 mm dekad⁻¹ and mean pvn_rrmse of 23.6%. Absolute errors (Figure 6b) were typically between 2–6 mm dekad⁻¹, while relative errors (Figure 6c) remained largely within 15–30%, except during isolated low-signal dry-season periods when pvn_rrmse temporarily exceeded 100%.

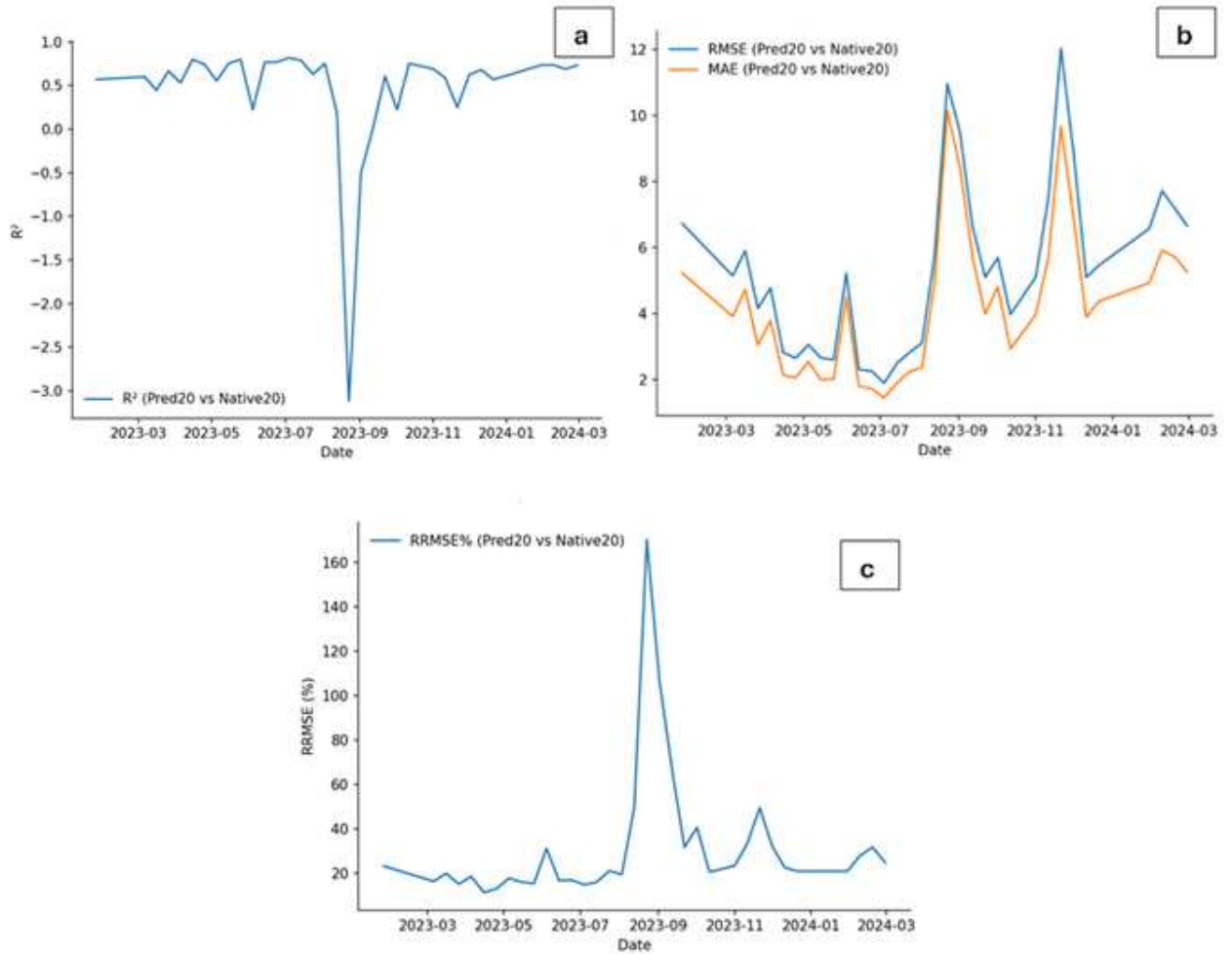


Figure 6 Temporal evolution of pixel-level (20 m) validation metrics for Baixo (2023–2024), showing (a) R^2 , (b) RMSE and MAE (mm dekad^{-1}), and (c) RRMSE (%), computed between predicted 20 m ET and native WaPOR Level-3 20 m ET for each dekad.

In Lamego (hold-out dekads outside the 21 January 2019–20 December 2020 model-development window; Figure 7), performance was comparable, with mean pvn_ r^2 of 0.62 (95% CI: 0.57–0.66; $n = 61$), mean pvn_rmse of $3.60 \text{ mm dekad}^{-1}$, and mean pvn_rrmse of 24.3%. Temporal patterns similarly indicate stable variance explanation across most dekads, with occasional short-lived degradations under low-ETa conditions.

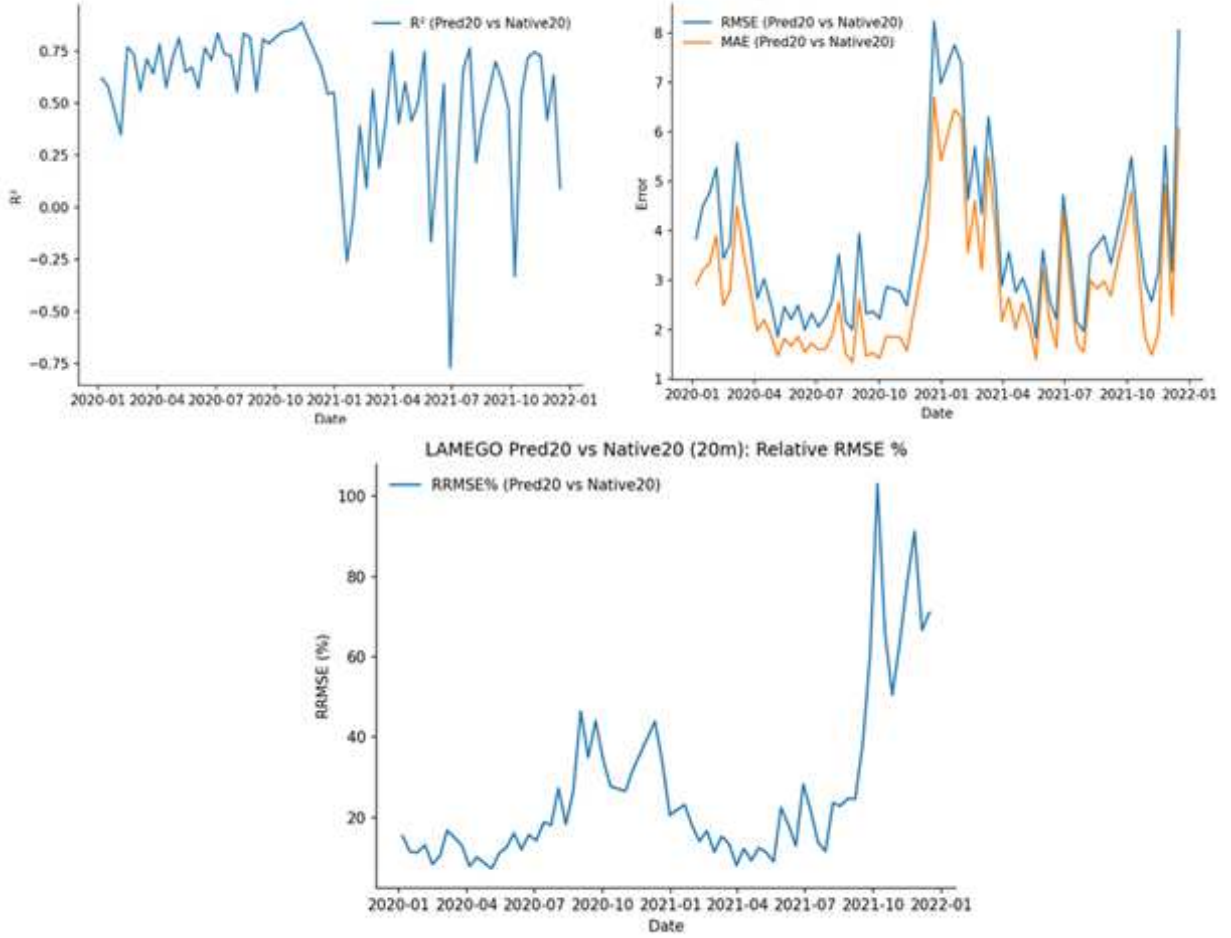


Figure 7. Pixel-level validation metrics for Lamego hold-out dekads outside the 21 January 2019–20 December 2020 model-development window. Shown are (a) R^2 , (b) RMSE and MAE (mm dekad^{-1}), and (c) RRMSE (%) between predicted and native 20 m ETa for each dekad.

Seasonal analysis showed significantly higher wet-season absolute errors at both sites ($p < 0.05$), whereas differences in pvn_r^2 and relative error were not statistically significant. Overall, the model consistently explained approximately 60–65% of pixel-scale spatial variability, demonstrating robust spatial fidelity under typical hydrological conditions, with instability confined to low-variance dry-season regimes rather than systematic structural bias.

3.3 Multi-Scale Aggregation (20 m $\rightarrow$ 300 m)

To evaluate large-scale consistency, predicted 20 m ETa (Pred20) and native 20 m ETa were aggregated to 300 m and compared with WaPOR Level-1 ETa (300 m) (Figures 8–9).

3.3.1 Baixo (2023–2024)

Across 33 dekads, Pred20 and Native20 showed highly similar aggregated error patterns (Figure 8). Bootstrap mean RMSE was $6.41 \text{ mm dekad}^{-1}$ (95% CI: 5.52–7.34) for Pred20 and

6.78 mm dekad⁻¹ (95% CI: 5.87–7.79) for Native20. Overlapping confidence intervals and non-significant paired tests ($p > 0.05$) indicate no large-scale degradation after downscaling.

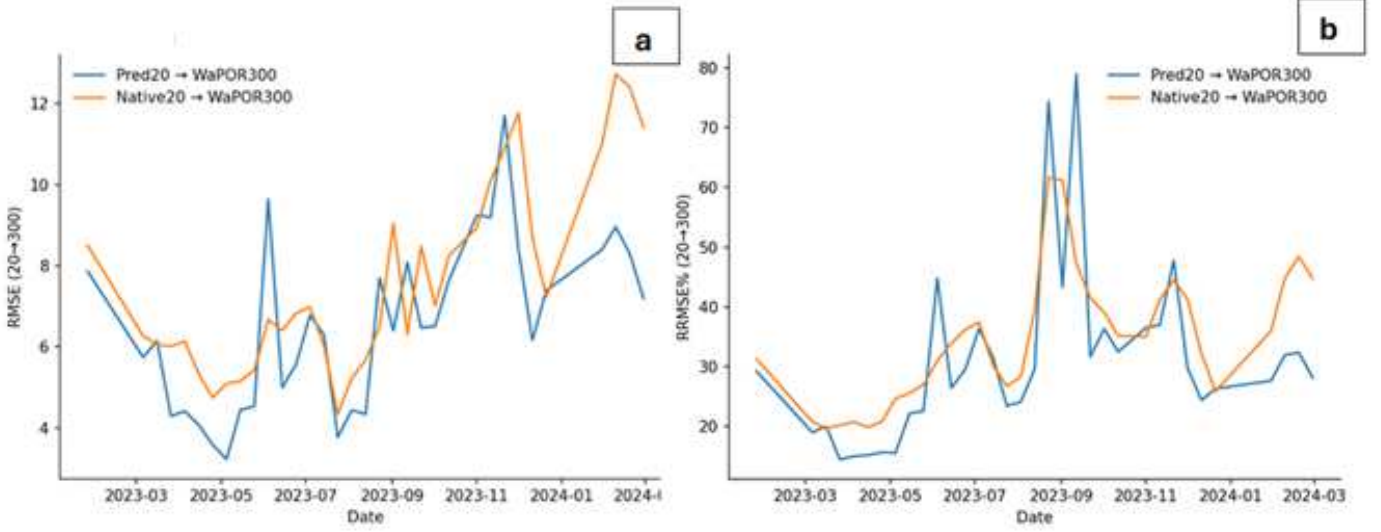


Figure 7 Baixo 20 m → 300 m aggregation results (2023–2024), showing RMSE and relative RMSE (%) between aggregated 20 m ET (Pred20 and Native20) and WaPOR 300 m ET. Pred20 closely follows Native20, indicating preserved large-scale consistency.

Seasonal increases in relative RMSE occurred in both products, reflecting hydrological variability rather than model bias.

3.3.2 Lamego (hold-out)

Similarly, across 68 dekads (Figure 9), Pred20 (5.12 mm dekad⁻¹; 95% CI: 4.55–5.71) and Native20 (5.35 mm dekad⁻¹; 95% CI: 4.79–5.95) showed no statistically significant difference in aggregated RMSE ($p > 0.05$). Relative errors increased during low-ETa periods for both datasets, again indicating signal effects rather than structural distortion.

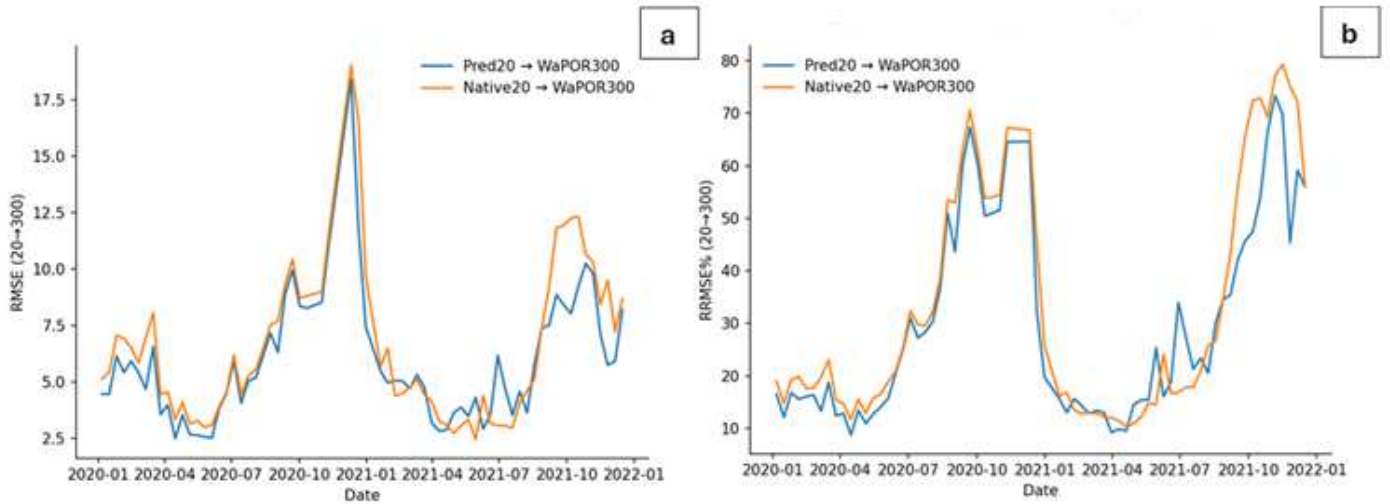


Figure 8 Dekadal large-scale consistency at Lamego hold-out dekads outside the 21 January 2019–20 December 2020 model-development window. (a) RMSE and (b) relative RMSE (RRMSE%) between aggregated Pred20 and Native20 ET (20 → 300 m) and WaPOR Level-1 ET. Similar error magnitudes indicate preservation of 300 m ET variability after downscaling

3.3.3 Cross-Site Synthesis

Across both sites, no statistically significant difference was detected between aggregated Pred20 and Native20 at 300 m. Confidence intervals overlap consistently, paired tests show no significant differences, and seasonal variability affects both products similarly. Combined with pixel-level results, this confirms that the 20 m framework enhances spatial detail with no detectable degradation of aggregated 300 m ETa magnitude.

4 Discussion

4.1 Performance of Dekadal ETa Downscaling in Fragmented Irrigation Landscapes

WaPOR-AETI-DownScaler v1.0 demonstrated strong performance in downscaling dekadal ETa from 300 m to 20 m across both study areas. Internal validation, used for model selection, achieved $R^2 = 0.883$, $RMSE = 3.95 \text{ mm dekad}^{-1}$, and $RRMSE = 19.5\%$, indicating that prediction errors were below one-fifth of mean ETa magnitude under within-domain conditions. Independent temporal hold-out testing, used for generalization assessment, yielded stable pixel-level explanatory power, with mean pvn_r^2 of approximately 0.62, absolute errors typically between 3–5 mm dekad^{-1} , and relative errors averaging approximately 23–24%.

The stability of relative RMSE across Baixo and Lamego indicates proportional consistency, meaning that errors scaled with ETa magnitude rather than reflecting systematic model bias. Although wet-season absolute errors increased with higher flux intensity, relative errors remained broadly stable. Negative pvn_r^2 values occurred only during very low-variance dry-season conditions and were attributed to weak spatial ETa signal rather than structural model failure. This behaviour is expected in low-signal periods, where small absolute deviations can produce unstable variance-based metrics.

4.2 Comparison with Existing Literature

The performance of WaPOR-AETI-DownScaler v1.0 is consistent with, and in some respects competitive with, previous ETa downscaling studies. In this study, internal validation achieved $R^2 = 0.883$, $\text{RMSE} = 3.95 \text{ mm dekad}^{-1}$, and $\text{RRMSE} = 19.5\%$, while independent temporal testing produced a mean pixel-level $\text{pvn_r}^2 \approx 0.62$ and $\text{RRMSE} \approx 23\text{--}24\%$. These values fall within the range reported by previous MODIS–Landsat and machine-learning-based ETa downscaling studies.

Ke et al. (2016) downscaled MODIS 8-day 1 km ET to 30 m using Landsat 8 predictors and machine-learning models, including Random Forest, Cubist and Support Vector Regression. Their Random Forest model produced the lowest error, with relative RMSE below 20% during model evaluation, while the final downscaled product had an average relative RMSE of 22% and an in situ validation $R^2 = 0.56$. The independent RRMSE obtained in the present study ($\sim 23\text{--}24\%$) is therefore comparable, although our framework operates at dekadal resolution and targets WaPOR 300 m to 20 m downscaling rather than MODIS 1 km to Landsat 30 m sharpening.

Singh et al. (2014) downscaled MODIS-based monthly actual evapotranspiration to Landsat scale over the Colorado River Basin using a MODIS–Landsat fusion approach. Their comparison with flux-site observations produced R^2 values ranging from 0.52 to 0.88, with monthly bias ranging from -16 to 22 mm month^{-1} . The independent pvn_r^2 of approximately 0.62 obtained here lies within this reported range, while the dekadal RMSE of $3\text{--}5 \text{ mm dekad}^{-1}$ suggests acceptable error magnitude for operational irrigation monitoring.

Deep-learning ETa downscaling studies have reported somewhat higher accuracy under controlled validation settings. Che et al. (2022) downscaled MODIS 8-day ET from 500 m to

30 m using a deep neural network trained with Landsat 8 reflectance, temperature and AgERA5 meteorological variables. Their DNN achieved average $R^2 = 0.67$, $\text{RMSD} = 2.63 \text{ mm } 8\text{-day}^{-1}$, and relative $\text{RMSD} = 14.25\%$, slightly outperforming Random Forest ($R^2 = 0.64$, $\text{RMSD} = 2.76 \text{ mm } 8\text{-day}^{-1}$, relative $\text{RMSD} = 14.92\%$). Compared with these values, WaPOR-AETI-DownScaler v1.0 shows similar explanatory power under independent testing ($\text{pvn_r}^2 \approx 0.62$) but higher proportional error ($\sim 23\text{--}24\%$), which is expected given the fragmented irrigation context, dekadal temporal scale, and reliance on WaPOR Level-3 as the reference target rather than flux-tower observations.

Overall, the comparison indicates that the present framework performs within the range of established ETa downscaling approaches. Its main contribution is not simply achieving higher pixel-level accuracy, but combining field-scale spatial refinement, dekadal operational timing, and multi-scale magnitude preservation. Unlike many previous studies that focus primarily on pixel-level R^2 or RMSE, this study explicitly tested whether aggregated 20 m predictions remained statistically consistent with WaPOR Level-1 ETa at 300 m. This conservation test strengthens the operational relevance of the framework for irrigation monitoring and basin-scale water accounting.

4.3 Drivers of Evapotranspiration Variability

Feature-importance analysis shows that spatial ETa variability was mainly controlled by vegetation moisture status, vegetation density, and seasonal dynamics. The high ranking of NDMI, NDVI, and FVC indicates that canopy water content, greenness, and fractional vegetation cover were key determinants of field-scale evapotranspiration patterns. This is physically consistent with the role of vegetation indices in representing photosynthetic activity, canopy density, and vegetation water status (Rouse et al. 1974; Gao 1996; Carlson & Ripley 1997; Huete et al. 2002). NDVI is widely used to quantify vegetation greenness and density, while NDMI/NDWI-type indices are sensitive to vegetation liquid water content, making them directly relevant for ETa estimation in irrigated landscapes.

The importance of FVC further supports the interpretation that ETa is strongly linked to the proportion of actively transpiring vegetation. Previous studies have used NDVI-derived FVC to separate vegetation transpiration from soil evaporation components in actual evapotranspiration estimation, demonstrating the relevance of vegetation cover information for ETa modelling (Carlson & Ripley 1997; Maselli et al. 2020).

The strong contribution of the harmonic seasonal variables, SIN_DOY and COS_DOY, further indicates that phenological timing and seasonal climate forcing structured ETa dynamics at dekadal scale. These variables helped the model distinguish between wet-season growth periods, dry-season irrigation signals, and low-activity periods when ETa variability was weaker. This is consistent with remote-sensing time-series studies that use seasonal or harmonic terms to capture vegetation phenology and temporal periodicity in land-surface processes.

ETa300m also ranked among the most important predictors, confirming the hybrid nature of the downscaling framework. In this structure, coarse-resolution WaPOR ETa provides the large-scale magnitude constraint, while high-resolution vegetation, rainfall, and terrain predictors redistribute that signal spatially to 20 m. Similar hybrid disaggregation approaches have been used in thermal sharpening and ET downscaling studies, where coarse-resolution thermal or ET information is sharpened using finer-resolution optical and vegetation indicators (Gao et al. 2012; Anderson et al. 2018; Ke et al. 2016; Che et al. 2022).

Overall, the feature-importance results support the physical plausibility of the model: ETa is primarily explained by vegetation condition, seasonal timing, and coarse-scale atmospheric–hydrological forcing, rather than by raw spectral reflectance alone. This interpretation is consistent with previous ET downscaling studies showing that vegetation indices, moisture indicators, land-surface temperature, terrain, and meteorological variables are key predictors of spatial ET variability (Ke et al. 2016; Che et al. 2022).

4.4 Seasonal and Low-Variance Regimes

Wet-season absolute errors were higher at both sites, reflecting greater ETa magnitude and stronger spatial heterogeneity during peak crop-growth periods. Similar seasonal error behaviour has been reported in satellite-based ETa evaluations, where performance varies with vegetation condition, climate forcing, land-surface heterogeneity and validation scale (Blatchford et al. 2020; Pan et al. 2020). However, relative RMSE remained stable, indicating that the increase in RMSE was mainly due to higher ETa fluxes rather than reduced model proportionality.

Negative $p_{vn_r^2}$ values occurred only during very low-variance dry-season dekads, when spatial ETa variability was close to zero. Under these conditions, small absolute errors can

produce unstable R^2 values because there is little spatial variance to explain. This behaviour is consistent with known limitations of correlation- and variance-based performance metrics in low-signal model evaluation, as well as challenges reported in multi-scale remote-sensing downscaling studies (Willmott 1982; Peng et al. 2017; Yang et al. 2020). After excluding extreme low-signal dekads, explanatory power remained stable at approximately 0.62, confirming that performance degradation was episodic rather than systematic.

4.5 Multi-Scale Consistency and Water Balance Integrity

A key contribution of this study is the explicit verification of multi-scale conservation. No statistically significant difference was detected between aggregated Pred20 ETa and Native20 ETa when compared with WaPOR Level-1 ETa at 300 m ($p > 0.05$), with overlapping bootstrap confidence intervals across sites. This indicates no detectable degradation of aggregated 300 m ETa magnitude after downscaling, which is essential for using downscaled ETa in basin-scale water accounting. Similar concerns have been highlighted in thermal sharpening, ET disaggregation, and remote-sensing data-fusion studies, where fine-resolution spatial detail must remain consistent with the parent coarse-resolution signal (Gao et al. 2012; Anderson et al. 2018; Ke et al. 2016; Che et al. 2022).

Such consistency between aggregated 20 m predictions and the parent 300 m signal is particularly important for operational water accounting, where ETa estimates are used to quantify water depletion, irrigation performance, and basin-scale water consumption. The result therefore supports the use of WaPOR-AETI-DownScaler v1.0 within water accounting frameworks such as WaPOR and Water Accounting Plus, which rely on spatially consistent ETa information for decision support in complex river basins (FAO 2018; Karimi et al. 2013; Blatchford et al. 2020).

4.6 Operational Implications

By combining 20 m spatial detail with preserved 300 m magnitude consistency, the framework enables field-scale irrigation diagnostics without compromising basin-scale reporting. The downscaled ETa maps can support identification of spatial variability in water consumption, detection of underperforming or stressed fields, comparison of irrigation intensity across schemes and development of scheme-level water accounting indicators.

The dekadal temporal resolution is also operationally important. It aligns with WaPOR's reporting structure and is frequent enough to support seasonal irrigation monitoring, while remaining less sensitive to daily noise and cloud contamination than very high-frequency products. This makes the framework suitable for near-real-time monitoring in regions where field observations are sparse and irrigation records are incomplete.

The approach is particularly relevant for transboundary basins such as the Limpopo, where local irrigation decisions and basin-scale water allocation are closely linked. Field-scale ETa information can improve local diagnostics, while preserved aggregation consistency allows the same product to contribute to broader water accounting and reporting frameworks. This dual scale of application is one of the main practical strengths of the framework.

4.7 Limitations and Future Work

Several limitations should be acknowledged. First, the model has so far been tested only in two Mozambican WaPOR Level-3 pilot areas, and transferability to other climatic regions, crop types and irrigation systems remains to be tested. Although Baixo and Lamego provide useful contrasts in scale and heterogeneity, they do not represent the full diversity of irrigated landscapes across Africa.

Second, because the model was trained using WaPOR Level-3 pilot areas as the high-resolution reference, the resulting product should be interpreted as a spatial refinement of WaPOR ETa rather than an independently observed ETa estimate. Further validation with flux towers, lysimeters, or field-scale water-balance observations would strengthen physical evaluation.

Third, the framework currently relies on optical Sentinel-2 predictors, which can be affected by cloud cover and data gaps. Although dekadal compositing reduces this problem, persistent cloudiness over wet-season dekads can degrade Sentinel-2 composite quality and, in extreme cases, leave individual dekads without usable optical predictors; in such situations the framework currently skips the affected dekad rather than produce a low-quality estimate, which briefly interrupts the operational ETa time series. Future work could explore integration of radar data, thermal data, soil moisture indicators or data-fusion approaches to improve robustness under cloudy conditions and to gap-fill missing dekads operationally.

Fourth, performance degradation under very low-variance conditions highlights limitations of R^2 and related variance-based metrics. Future studies should explore alternative evaluation metrics for low-signal regimes, including bias, MAE, relative error and categorical performance indicators for irrigation activity. This would provide a more balanced interpretation of performance during dry-season or low-ETa periods.

Finally, future work should compare CatBoost with alternative machine-learning and deep-learning models, including Random Forest, XGBoost and neural-network architectures, under identical train–validation–test splits. Such comparisons should consider not only predictive accuracy, but also data requirements, computational cost, interpretability and operational robustness. For WaPOR-based applications in data-scarce regions, a scalable and interpretable approach may be more valuable than marginal improvements in accuracy.

5 Conclusion

WaPOR-AETI-DownScaler v1.0 provides a reproducible and operational framework for downscaling dekadal WaPOR actual evapotranspiration from approximately 300 m to 20 m resolution in fragmented irrigation landscapes. Internal validation, used for model selection, showed strong performance ($R^2 = 0.883$; $RMSE = 3.95 \text{ mm dekad}^{-1}$; $RRMSE = 19.5\%$), while independent temporal hold-out testing demonstrated stable generalization, with mean pixel-level R^2 of approximately 0.62 and relative RMSE of 23–24%. These results indicate that the framework can reproduce a substantial proportion of fine-scale spatial ETa variability under operational dekadal conditions.

Vegetation moisture and greenness indicators, particularly NDMI, NDVI and FVC, together with seasonal harmonic variables, were the dominant controls of spatial ETa heterogeneity. Coarse-resolution WaPOR ETa also provided an important large-scale magnitude constraint, supporting the interpretation of the model as a spatial refinement framework rather than an independent ETa retrieval system. Aggregation of the 20 m predictions back to 300 m showed no statistically detectable degradation relative to the parent WaPOR Level-1 ETa signal, indicating that the model enhanced spatial detail while maintaining consistency with the original coarse-resolution magnitude.

Overall, the framework combines field-scale heterogeneity mapping with explicit multi-scale consistency assessment, making it suitable for irrigation monitoring, scheme-level diagnostics

and water-accounting applications where both spatial detail and large-scale consistency are required. However, because the model was evaluated in two Mozambican WaPOR Level-3 pilot areas, further testing across additional climates, crop systems and irrigation settings, as well as validation against independent field observations, is needed before wider regional deployment.

Acknowledgements

The authors acknowledge the data providers whose open-access datasets made this study possible, including the FAO WaPOR programme for evapotranspiration products, ESA Copernicus for Sentinel-2 and Copernicus DEM data, the Climate Hazards Center for CHIRPS rainfall data, and ECMWF for ERA5-Land meteorological data. We also acknowledge the FAO WaPOR team, particularly Virginie Gillet and Livia Peiser, for their support and technical engagement related to WaPOR data and applications.

Declarations

Funding

This work was funded by the Government of the Netherlands through the Food and Agriculture Organization of the United Nations (FAO), under FAO Project Symbol GCP/INT/729/NET and Budget Code TF.NSLDD.TFNL110020514. The associated administrative records include GRMS Supplier Number 38329 and PO Number 351633. Competing interests,

The authors have no relevant financial or non-financial interests to disclose.

Clinical trial number

Not applicable.

Ethics approval

Not applicable. This study did not involve human participants or animals; all data are remotely sensed or model-derived.

Consent to participate

Not applicable.

Consent for publication

Not applicable.

Data availability

All input datasets used in this study are openly available: WaPOR Level-1 and Level-3 ETa products from the FAO WaPOR portal (<https://wapor.apps.fao.org>); Sentinel-2 surface reflectance from the Copernicus Open Access Hub; CHIRPS precipitation from the Climate

Hazards Center (<https://www.chc.ucsb.edu/data/chirps>); ERA5-Land from the Copernicus Climate Data Store; Copernicus DEM GLO-30 from the European Space Agency (<https://doi.org/10.5270/ESA-c5d3d65>); ESA WorldCover 2021 v200 (<https://doi.org/10.5281/zenodo.7254221>). Trained model weights and per-dekad evaluation outputs are available from the corresponding author on reasonable request.

Code availability

The full implementation of the WaPOR-AETI-DownScaler v1.0 workflow is publicly available at <https://github.com/ZoloKiala/WaPOR-downscaling>. The trained model is also deployed as an operational web application at https://huggingface.co/spaces/IWMIHQ/WaPOR_20_m_Downscaler.

Author contributions

Zolo Kiala, Petra Schmitter, and Mariangel Garcia Andarcia contributed to the study conception and design. Material preparation, data processing, model development, analysis, and visualization were performed by Zolo Kiala. Petra Schmitter and Mariangel Garcia Andarcia contributed to the methodological framing, interpretation of results, and review of the manuscript. The first draft of the manuscript was written by Zolo Kiala. All authors commented on previous versions of the manuscript, read and approved the final manuscript, and agreed to be accountable for the work.

References

- Allen RG, Pereira LS, Raes D, Smith M (2007) Satellite-based energy balance for mapping evapotranspiration with internalized calibration (METRIC). *Journal of Irrigation and Drainage Engineering* 133(4):380–394. [https://doi.org/10.1061/\(ASCE\)0733-9437\(2007\)133:4\(380\)](https://doi.org/10.1061/(ASCE)0733-9437(2007)133:4(380))
- Anderson MC, Allen RG, Morse A, Kustas WP (2012) Use of Landsat thermal imagery in monitoring evapotranspiration and managing water resources. *Remote Sensing of Environment* 122:50–65. <https://doi.org/10.1016/j.rse.2011.08.025>
- Anderson MC, Gao F, Knipper K, Hain C, Dulaney W, Baldocchi D, Eichelmann E, Hemes K, Yang Y, Medellín-Azuara J et al (2018) Field-scale assessment of land and water use change over the California Delta using remote sensing. *Remote Sensing* 10(6):889. <https://doi.org/10.3390/rs10060889>
- Bastiaanssen WGM, Menenti M, Feddes RA, Holtslag AAM (1998) A remote sensing surface energy balance algorithm for land (SEBAL): 1. Formulation. *Journal of Hydrology* 212–213:198–212. [https://doi.org/10.1016/S0022-1694\(98\)00253-4](https://doi.org/10.1016/S0022-1694(98)00253-4)

Beer C, Reichstein M, Tomelleri E, Ciais P, Jung M, Carvalhais N, Rödenbeck C, Arain MA, Baldocchi D, Bonan GB, Bondeau A, Cescatti A, Lasslop G, Lindroth A, Lomas M, Luyssaert S, Margolis H, Oleson KW, Rouspard O, Veenendaal E, Viovy N, Williams C, Woodward FI, Papale D (2010) Terrestrial gross carbon dioxide uptake: Global distribution and covariation with climate. *Science* 329(5993):834–838. <https://doi.org/10.1126/science.1184984>

Biewald L (2020) Experiment tracking with Weights and Biases. Software available from wandb.com

Blatchford ML, Mannaerts CM, Njuki SM, Nouri H, Zeng Y, Pelgrum H, Wonink S, Karimi P (2020) Evaluation of WaPOR V2 evapotranspiration products across Africa. *Hydrological Processes* 34(15):3200–3221. <https://doi.org/10.1002/hyp.13791>

Carlson TN, Ripley DA (1997) On the relation between NDVI, fractional vegetation cover, and leaf area index. *Remote Sensing of Environment* 62(3):241–252. [https://doi.org/10.1016/S0034-4257\(97\)00104-1](https://doi.org/10.1016/S0034-4257(97)00104-1)

Che X, Zhang HK, Sun Q, Ouyang Z, Liu J (2022) MODIS evapotranspiration downscaling using a deep neural network trained using Landsat 8 reflectance and temperature data. *Remote Sensing* 14(22):5876. <https://doi.org/10.3390/rs14225876>

Copernicus DEM (2020) Copernicus Digital Elevation Model (DEM) GLO-30 [Data set]. European Space Agency. <https://doi.org/10.5270/ESA-c5d3d65>

Drusch M, Del Bello U, Carlier S, Colin O, Fernandez V, Gascon F, Hoersch B, Isola C, Laberinti P, Martimort P, Meygret A, Spoto F, Sy O, Marchese F, Bargellini P (2012) Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sensing of Environment* 120:25–36. <https://doi.org/10.1016/j.rse.2011.11.026>

FAO (2017) The future of food and agriculture – Trends and challenges. Food and Agriculture Organization of the United Nations

FAO (2018) WaPOR Database Methodology: Level 1, 2 and 3. Food and Agriculture Organization of the United Nations

Fisher JB, Hook SJ, Allen RG, Anderson MC, French AN, Hain CR, Hulley GC, Wood EF, Yebra M (2020) ECOSTRESS: NASA's next generation mission to measure evapotranspiration from the International Space Station. *Water Resources Research* 56(4):e2019WR026058. <https://doi.org/10.1029/2019WR026058>

- Funk CC, Peterson PJ, Landsfeld MF, Pedreros DH, Verdin JP, Shukla S, Husak GJ, Rowland JD, Harrison L, Hoell A, Michaelsen JC (2015) The climate hazards infrared precipitation with stations—A new environmental record for monitoring extremes. *Scientific Data* 2:150066. <https://doi.org/10.1038/sdata.2015.66>
- Gao B-C (1996) NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment* 58(3):257–266. [https://doi.org/10.1016/S0034-4257\(96\)00067-3](https://doi.org/10.1016/S0034-4257(96)00067-3)
- Gao F, Kustas WP, Anderson MC (2012) A data mining approach for sharpening thermal satellite imagery over land. *Remote Sensing* 4(11):3287–3319. <https://doi.org/10.3390/rs4113287>
- Huete AR (1988) A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment* 25(3):295–309. [https://doi.org/10.1016/0034-4257\(88\)90106-X](https://doi.org/10.1016/0034-4257(88)90106-X)
- Huete A, Didan K, Miura T, Rodriguez EP, Gao X, Ferreira LG (2002) Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment* 83(1–2):195–213. [https://doi.org/10.1016/S0034-4257\(02\)00096-2](https://doi.org/10.1016/S0034-4257(02)00096-2)
- Karimi P, Bastiaanssen WGM, Molden D (2013) Water Accounting Plus (WA+) – A water accounting procedure for complex river basins based on satellite measurements. *Hydrology and Earth System Sciences* 17:2459–2472. <https://doi.org/10.5194/hess-17-2459-2013>
- Ke Y, Im J, Park S, Gong H (2016) Downscaling of MODIS one kilometer evapotranspiration using Landsat-8 data and machine learning approaches. *Remote Sensing* 8(3):215. <https://doi.org/10.3390/rs8030215>
- Key CH, Benson NC (2006) Landscape Assessment (LA): Sampling and analysis methods. In FIREMON: Fire Effects Monitoring and Inventory System (pp. LA-1–LA-55). Gen. Tech. Rep. RMRS-GTR-164-CD. USDA Forest Service, Rocky Mountain Research Station
- Louis J, Debaecker V, Pflug B, Main-Knorn M, Bieniarz J, Müller-Wilm U, Cadau E, Gascon F (2016) Sentinel-2 Sen2Cor: L2A processor for users. In Living Planet Symposium 2016, ESA SP-740
- Martens B, Miralles DG, Lievens H, van der Schalie R, de Jeu RAM, Fernández-Prieto D, Beck HE, Dorigo WA, Verhoest NEC (2017) GLEAM v3: Satellite-based land evaporation and root-

zone soil moisture. *Geoscientific Model Development* 10(5):1903–1925.
<https://doi.org/10.5194/gmd-10-1903-2017>

Maselli F, Chiesi M, Angeli L, Fibbi L, Rapi B, Romani M, Sabatini F, Battista P (2020) An improved NDVI-based method to predict actual evapotranspiration of irrigated grasses and crops. *Agricultural Water Management* 233:106077.
<https://doi.org/10.1016/j.agwat.2020.106077>

Molden D, Oweis T, Steduto P, Kijne J, Hanjra M, Bindraban P (2010) Improving agricultural water productivity: Between optimism and caution. *Agricultural Water Management* 97(4):528–535. <https://doi.org/10.1016/j.agwat.2009.03.023>

Mu Q, Zhao M, Running SW (2011) Improvements to a MODIS global terrestrial evapotranspiration algorithm. *Remote Sensing of Environment* 115(8):1781–1800.
<https://doi.org/10.1016/j.rse.2011.02.019>

Muñoz-Sabater J, Dutra E, Agustí-Panareda A, Albergel C, Arduini G, Balsamo G, Boussetta S, Choulga M, Harrigan S, Hersbach H, Martens B, Miralles DG, Piles M, Rodríguez-Fernández NJ, Zsoter E, Buontempo C, Thépaut JN (2021) ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. *Earth System Science Data* 13(9):4349–4383.
<https://doi.org/10.5194/essd-13-4349-2021>

Pan S, Pan N, Tian H, Friedlingstein P, Sitch S, Shi H, Arora VK, Haverd V, Jain AK, Kato E, Lienert S, Lombardozzi D, Nabel JEMS, Ottlé C, Poulter B, Zaehle S, Running SW (2020) Evaluation of global terrestrial evapotranspiration using state-of-the-art approaches in remote sensing, machine learning and land surface modeling. *Hydrology and Earth System Sciences* 24:1485–1509. <https://doi.org/10.5194/hess-24-1485-2020>

Peng J, Loew A, Merlin O, Verhoest NEC (2017) A review of spatial downscaling of satellite remotely sensed soil moisture. *Reviews of Geophysics* 55(2):341–366.
<https://doi.org/10.1002/2016RG000543>

Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A (2018) CatBoost: Unbiased boosting with categorical features. *Advances in Neural Information Processing Systems* 31:6638–6648

Rouse JW, Haas RH, Schell JA, Deering DW (1974) Monitoring vegetation systems in the Great Plains with ERTS. In Third Earth Resources Technology Satellite (ERTS) Symposium (pp. 309–317)

Senay GB, Bohms S, Singh RK, Gowda PH, Velpuri NM, Alemu H, Verdin JP (2013) Operational evapotranspiration mapping using remote sensing and weather datasets: A new parameterization for the SSEB approach. *Journal of the American Water Resources Association* 49(3):577–591. <https://doi.org/10.1111/jawr.12057>

Singh RK, Senay GB, Velpuri NM, Bohms S, Verdin JP (2014) On the downscaling of actual evapotranspiration maps based on combination of MODIS and Landsat-based actual evapotranspiration estimates. *Remote Sensing* 6(11):10483–10509. <https://doi.org/10.3390/rs61110483>

Sun X, Zhan W, Xu H (2019) Downscaling of MODIS evapotranspiration using land surface temperature and NDVI in a heterogeneous agricultural region. *Remote Sensing* 11(3):294. <https://doi.org/10.3390/rs11030294>

Willmott CJ (1982) Some comments on the evaluation of model performance. *Bulletin of the American Meteorological Society* 63(11):1309–1313. [https://doi.org/10.1175/1520-0477\(1982\)063<1309:SCOTEO>2.0.CO;2](https://doi.org/10.1175/1520-0477(1982)063<1309:SCOTEO>2.0.CO;2)

Wilson JP, Gallant JC (eds) (2000) *Terrain analysis: Principles and applications*. John Wiley & Sons

WWAP (2021) *The United Nations World Water Development Report 2021: Valuing Water*. UNESCO World Water Assessment Programme

Xu J, Singh VP, Paredes P (2018) A review of evapotranspiration downscaling methods using remote sensing. *Journal of Hydrology* 566:173–193. <https://doi.org/10.1016/j.jhydrol.2018.09.038>

Yang Y, Long D, Guan H et al (2020) Investigating impacts of drought and disturbance on evapotranspiration over global forests in recent two decades. *Remote Sensing of Environment* 248:111415. <https://doi.org/10.1016/j.rse.2018.12.017>

You L, Ringler C, Wood-Sichra U, Robertson R, Wood S, Zhu T, Nelson G, Guo Z, Sun Y (2011) What is the irrigation potential for Africa? A combined biophysical and socioeconomic approach. *Food Policy* 36(6):770–782. <https://doi.org/10.1016/j.foodpol.2011.09.001>

Xie H, You L, Wielgosz B, Ringler C (2014) Estimating the potential for expanding smallholder irrigation in Sub-Saharan Africa. *Agricultural Water Management* 131:183–193. <https://doi.org/10.1016/j.agwat.2013.08.011>

Zanaga D, Van De Kerchove R, Daems D, De Keersmaecker W, Brockmann C, Kirches G, Wevers J, Cartus O, Santoro M, Fritz S, Lesiv M, Herold M, Tsendbazar N-E, Xu P, Ramoino F, Arino O (2022) ESA WorldCover 10 m 2021 v200 [Data set]. European Space Agency. <https://doi.org/10.5281/zenodo.7254221>