Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

AlphaEarth and evapotranspiration prediction in Jordan

Samer Abaddi

eng.sameralabbadi@gmail.com

IDentity Research https://orcid.org/0000-0003-4612-7384

Research Article

Keywords: AlphaEarth, WaPOR, evapotranspiration, geospatial foundation models, spatial transfer, Jordan

Posted Date: October 29th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11417788/v1

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

Additional Declarations: The authors declare no competing interests.

AlphaEarth and evapotranspiration prediction in Jordan

Samer Abaddi
IDentity Research, Amman, Jordan
ORCID: https://orcid.org/0000-0003-4612-7384
Correspondence: Samer Abaddi, IDentity Research, Amman, Jordan. Email: eng.sameralabbadi@gmail.com

Abstract

The value of geospatial foundation embeddings for monthly prediction depends on whether they improve on information already available from conventional predictors. A retrospective paired study evaluated prior-year AlphaEarth embeddings for next-month WaPOR actual evapotranspiration and interception in the Jordan Valley and Mafraq. The cohort comprised 352 cropland-screened cells in 44 blocks. Models were developed using 2019–2023 data and held fixed for evaluation on 8,448 cell-months in 2024–2025. Under an internally locked protocol, conventional and embedding-augmented histogram gradient boosting models were compared using region-balanced and block-balanced mean absolute error (MAE), with ridge regression, persistence and seasonal naive as comparators. Four-fold spatial evaluation with a start-of-season 5 km buffer and bidirectional regional transfer was included; prior local target history remained available throughout. All predictions were sealed before scoring. Temporal MAE increased from 6.430 to 6.585 mm/month with augmentation, a 2.41% deterioration. Conventional boosting reduced MAE by 25.96% relative to persistence and 32.32% relative to seasonal naive. Ridge also deteriorated, and pooled buffered spatial MAE increased slightly from 7.059 to 7.079 mm/month. Augmentation reduced transferred-model MAE by 6.55% into the Jordan Valley and 5.40% into Mafraq; only transfer into the Jordan Valley outperformed both local-history baselines. Embedding value was therefore conditional on the prediction setting and comparator. These results concern reproduction of satellite retrievals.

Independent physical validation and verification of historical product availability are needed before operational water-management use can be supported.

Keywords: AlphaEarth; WaPOR; evapotranspiration; geospatial foundation models; spatial transfer; Jordan

Introduction

Nearly three quarters of global freshwater withdrawals—approximately 72%—are used by agriculture, placing water information at the center of food security (United Nations Educational, Scientific and Cultural Organization [UNESCO], 2024). As competing demands intensify, decisions about agricultural production must be informed by where water is used and how that use changes through time. Withdrawal statistics establish the magnitude of the challenge; spatially explicit information is also required to connect resource pressures with the landscapes in which water is consumed.

This requirement is particularly consequential in the Middle East and North Africa (MENA). In a World Bank assessment, regional annual water availability per person was projected to fall below the absolute-scarcity threshold of 500 m^3 by 2030 (de Waal et al., 2023). The problem was attributed to interacting pressures from climate change, population growth, and socioeconomic development, while the limits of infrastructure expansion without institutional reform were emphasized. Reliable environmental information is therefore needed alongside allocation and governance measures, so that changes in agricultural water use can be assessed within explicit resource constraints.

Within MENA, the Gulf Cooperation Council (GCC) illustrates the difficulty of responding primarily through additional supply. Renewable freshwater availability was reported to be below 100 m^3 per person annually in many GCC countries, with substantial dependence on non-renewable groundwater and energy-intensive desalination (Chattha et al., 2024). In the same assessment, water efficiency, wastewater reuse, and renewable energy for desalination were identified as complementary priorities. The Gulf experience underscores why technological capacity to produce water must be accompanied by evidence about its use; regional scarcity cannot be understood through supply volumes alone.

In Jordan, these pressures are expressed under an exceptionally restricted renewable resource base. Annual renewable freshwater availability was estimated at 61 m^3 per person in

2021, far below the 500 m³ absolute-scarcity threshold (Ministry of Water and Irrigation [MWI], 2023). Approximately 92% of the country receives less than 200 mm of rainfall annually. Moreover, about 450 million m³ was abstracted from renewable groundwater resources in 2021, compared with an estimated annual safe yield of 280 million m³ (MWI, 2023, pp. 22, 25). These figures jointly demonstrate why agricultural water assessments must account for environmental variation and sustained pressure on existing resources.

Satellite observations provide a practical basis for characterizing evapotranspiration (ET) consistently across space. Through the WaPOR platform, agricultural water productivity can be monitored using openly accessible remotely sensed data (Food and Agriculture Organization of the United Nations [FAO], 2026a). Actual evapotranspiration and interception (AETI) provides a record of water transferred to the atmosphere. Its relevance in Jordan has been demonstrated through water accounting in the North Jordan Valley, where WaPOR was assessed alongside climatic observations, field information, and calibrated energy-balance estimates (Al-Bakri et al., 2022). However, uncertainty in satellite retrievals was also shown to affect accounting outputs. Predictive agreement with WaPOR must consequently be distinguished from validation against independently measured water fluxes.

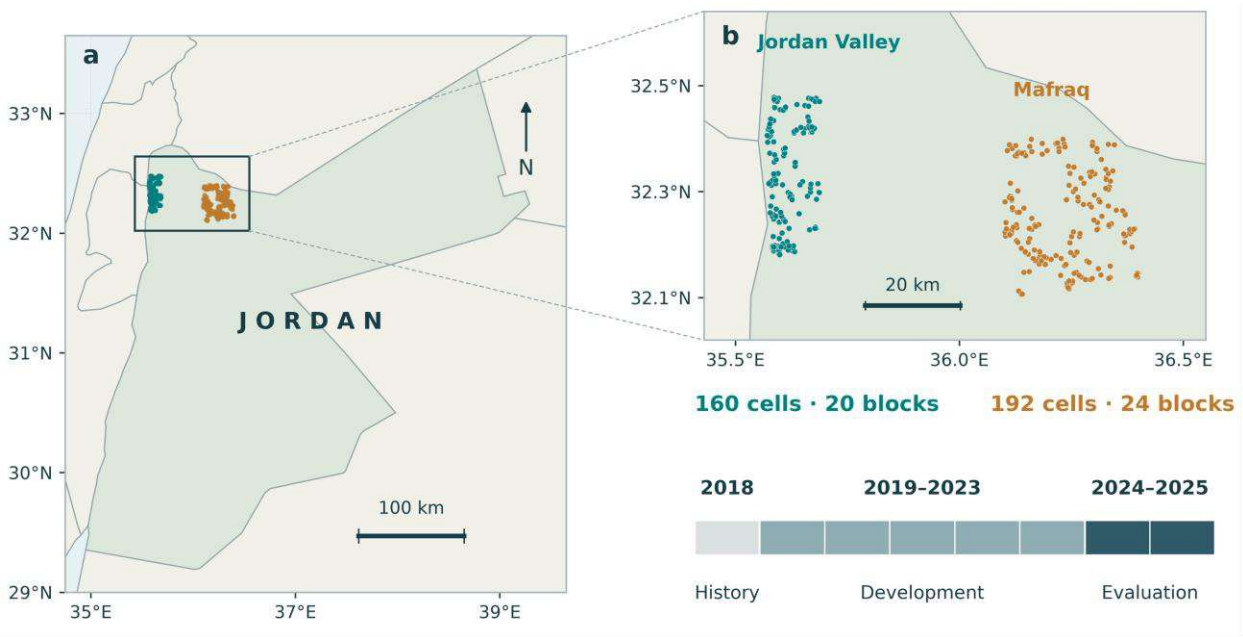
Two agricultural sampling areas in northern Jordan are considered here: the Jordan Valley and Mafraq. The study population comprises 352 cropland-screened cells grouped into 44 spatial blocks. Their locations and the temporal organization of the study are shown in Figure 1. This geographically bounded setting permits the value of additional satellite information to be examined both within sampled locations and across separated agricultural landscapes.

Geospatial foundation models offer a further source of environmental information. In the AlphaEarth Foundations preprint, heterogeneous Earth-observation inputs were represented through compact embeddings that could be used in downstream mapping tasks without retraining the foundation model (Brown et al., 2025). Annual embeddings provide a means of describing surface characteristics that may be incompletely represented by individual vegetation indices or terrain variables. Monthly OpenET proxy prediction was already included in the original

evaluation. The scientific opportunity is therefore more specific than applying a new representation to an ET-related target: its additional information must be tested against predictors that already describe seasonality, antecedent water use, weather, and vegetation.

Figure 1. Jordan Study Locations and the Temporal Structure of the Evaluation

Note. (a) National context and the extent enlarged in panel b. (b) The original 352 sampled cell centers: 160 in the Jordan Valley and 192 in Mafraq, grouped into 20 and 24 blocks, respectively. Coordinates are WGS84; scale bars are calculated geodesically at their displayed latitudes. Basemap: Natural Earth (2026), 1:50 million, version 5.1.1, public domain.



The remaining uncertainty concerns the conditions under which that additional information is useful. Information encoded in an annual representation may overlap with local historical records, while its contribution could differ when predictions are transferred between landscapes. Evidence from catchment modeling supports treating temporal and spatial generalization as distinct questions (Liu et al., 2026). For the Jordanian sampling areas, the gap addressed here is whether prior-year AlphaEarth embeddings contribute incremental information for next-month AETI prediction after conventional environmental and historical predictors have been included. Neither broad mapping performance nor evidence from a different hydrological target resolves this particular comparison.

The credibility of such a comparison depends on its evaluation design. Predictive errors can be underestimated when spatial or temporal dependence is ignored, and blocking strategies must be aligned with the intended application (Roberts et al., 2017). Future-month prediction at previously sampled cells, prediction in held-out spatial blocks, and transfer between regions therefore require separate assessments. Comparable rows and training-only preprocessing are also necessary to prevent differences in data availability from being mistaken for representation value. Because neighboring cells can share weather inputs, spatial buffers should be treated as controls on proximity.

The primary objective is to quantify whether the previous year's AlphaEarth embeddings improve sequential, one-month-ahead prediction of WaPOR AETI beyond calendar, AETI history, weather, terrain, and verified vegetation predictors. Secondary objectives are to examine transfer across held-out blocks and between the two sampling areas, and to assess sensitivity to predefined vegetation-support and land-cover screens. Model development is restricted to 2019–2023, with 2024–2025 reserved for evaluation. The experiment is retrospective: observation-time restrictions are applied, while historical product-release availability is not assumed.

Three contributions are made through this design:

- This study is a benchmark for Jordan since paired conventional and embedding-augmented models are compared on identical cells and months, with persistence and seasonal-naive predictions retained as practical reference points.
- A distinct test of generalization through temporal holdout, four spatial folds with 5 km buffers, and both regional transfer directions are assessed separately; prior local AETI remains available at held-out locations.
- An auditable assessment since missing-data rules and preprocessing are fixed within training, and frozen-model predictions are sealed before scoring, supporting transparent comparison without presuming improvement from the embeddings.

Literature review

Satellite estimates and agricultural water accounting

Satellite evapotranspiration (ET) products provide spatial evidence for agricultural water accounting, but their usefulness depends on the distinction between estimating physical water fluxes and reproducing an existing retrieval. This distinction also governs how geospatial foundation models should be evaluated. The literature reviewed here connects satellite-product validation, agricultural artificial intelligence, and representation transfer to assess whether AlphaEarth embeddings contribute information beyond conventional predictors in Jordan. The relevant question is incremental predictive value under specified temporal and spatial conditions, rather than accuracy alone.

According to Blatchford et al. (2020), WaPOR V2 captured broad spatiotemporal ET patterns across Africa, while comparisons with 14 eddy-covariance stations revealed variable performance and overestimation in irrigated areas. Complementing this evidence, Volk et al. (2024) assessed OpenET against 152 in situ stations and reported a cropland ensemble mean absolute error (MAE) of 15.8 mm/month.

Al-Bakri et al. (2022) studied the North Jordan Valley, combining WaPOR with climatic and ground information, Sentinel-2 and Landsat 8 imagery, and calibrated Surface Energy Balance Algorithm for Land estimates. WaPOR overestimation of ET in irrigated areas affected estimated recoverable water. Their contribution establishes both the management relevance of satellite observations and the consequences of retrieval uncertainty in Jordan. It leaves a different question open: whether learned representations improve subsequent monthly prediction after local ET history, weather, terrain, seasonality, and vegetation information are already available.

Agricultural intelligence and transferable representations

At the application level, Abaddi's (2025) Mush-Room study integrated Internet of Things (IoT) sensors, pretrained image-classification networks, and GPT-4o within an agricultural prototype. It illustrates how heterogeneous observations and learned representations can be

combined with a user-facing decision interface. Brown et al. (2025) introduced AlphaEarth Foundations, presenting compact representations learned from multiple Earth-observation sources. Crucially, their benchmarks already included monthly OpenET ensemble prediction. Separate evidence from hydrology suggests that this added value may depend on the generalization task. Liu et al. (2026) found similar temporal performance using AlphaEarth versus traditional catchment attributes across 531 basins: median Nash–Sutcliffe efficiencies were 0.779 and 0.781, respectively. Combining embeddings with traditional attributes improved the corresponding score from 0.683 to 0.720 in a separate global ungauged-basin experiment. Finally, in a cropland-mapping study, Mughees et al. (2026) reported 93.7% overall accuracy against held-out reference patches and assessed predictions independently through human interpretation. Their one-block validation exposed differences between agreement with an existing map and agreement with independent interpreters.

Evaluation design and the contribution of the Jordan study

Roberts et al. (2017) explain why random cross-validation can underestimate prediction error in structured environmental data. Blocking must reflect the intended prediction setting; excessively restrictive splits can instead measure extrapolation when interpolation is the objective. For Jordan, temporal holdout, buffered spatial holdout, and leave-region-out transfer therefore answer separate questions. A buffer is a design choice, particularly when locations share coarse weather grids. Likewise, restricting embedding years does not establish historical availability of retrospectively revised products.

The present study addresses these distinctions through a fixed comparison of conventional predictors with and without 64 prior-year AlphaEarth dimensions. Development used 2019–2023; frozen predictions were sealed before scoring 2024–2025. Identical comparison rows, training-only missing-data treatment, persistence and seasonal-naive baselines, and separate spatial evaluations make the added-value claim auditable. Held-out locations retain prior

local actual ET and interception observations, so the spatial tests are not ungauged-site experiments. Table 1 positions this contribution against the reviewed evidence.

Table 1. Satellite Evidence and the Contribution of the Jordan Prediction Study

Study	Established evidence	Relation to the present study
Blatchford et al. (2020); Volk et al. (2024)	Validated satellite ET against independent observations; accuracy varied across settings.	Jordan predictions assess WaPOR fidelity; independent field validation remains necessary.
Al-Bakri et al. (2022)	Linked WaPOR retrieval uncertainty to North Jordan Valley water accounting.	Tests added predictive value after conventional environmental and historical inputs.
Abaddi (2025)	Integrated sensors, image classifiers, and GPT-4o in an agricultural prototype.	Extends agricultural AI integration to an audited geospatial comparison.
Brown et al. (2025)	Introduced AlphaEarth and demonstrated OpenET proxy prediction.	Tests incremental next-month value rather than claiming the first ET application.
Liu et al. (2026)	Found comparable temporal skill and stronger ungauged-basin transfer with combined descriptors.	Separates temporal, buffered spatial, and two directional regional tests.
Mughees et al. (2026)	Demonstrated cropland mapping and a local independent human audit.	Keeps reference agreement distinct from physical ET validation.

Note. Brown et al. (2025) and Mughees et al. (2026) are preprints. Present-study results concern the frozen histogram gradient boosting comparisons under protocol v1. MAE = mean absolute error; ET = evapotranspiration.

The resulting contribution is a bounded empirical test of when embeddings help: the primary temporal comparison deteriorated, while transfer gains depended on destination and comparator. This evidence supports further independent validation and targeted transfer research,

without establishing field ET accuracy, causal irrigation effects, or universal superiority of foundation embeddings.

Methodology

Study design

A retrospective, paired prediction study was conducted to assess the incremental information supplied by prior-year AlphaEarth embeddings for next-month WaPOR actual evapotranspiration and interception (AETI). Protocol v1 was internally hash-locked on 16 October 2026 after source and covariate quality review, before model fitting and evaluation scoring; it was not externally preregistered. Conventional and embedding-augmented models were compared on identical observations. The workflow is summarized in Figure 2.

The primary task was sequential prediction at previously sampled cells. Secondary tasks concerned held-out spatial blocks and transfer between the Jordan Valley and Mafraq sampling areas. Strictly prior local AETI remained available in all tasks; consequently, spatial transfer did not represent prediction at ungauged locations. The target was a satellite retrieval, not field-measured evapotranspiration, irrigation demand, or crop stress. Product-release timing and foundation-model pretraining cutoffs were not established, precluding claims of historically operational forecasting.

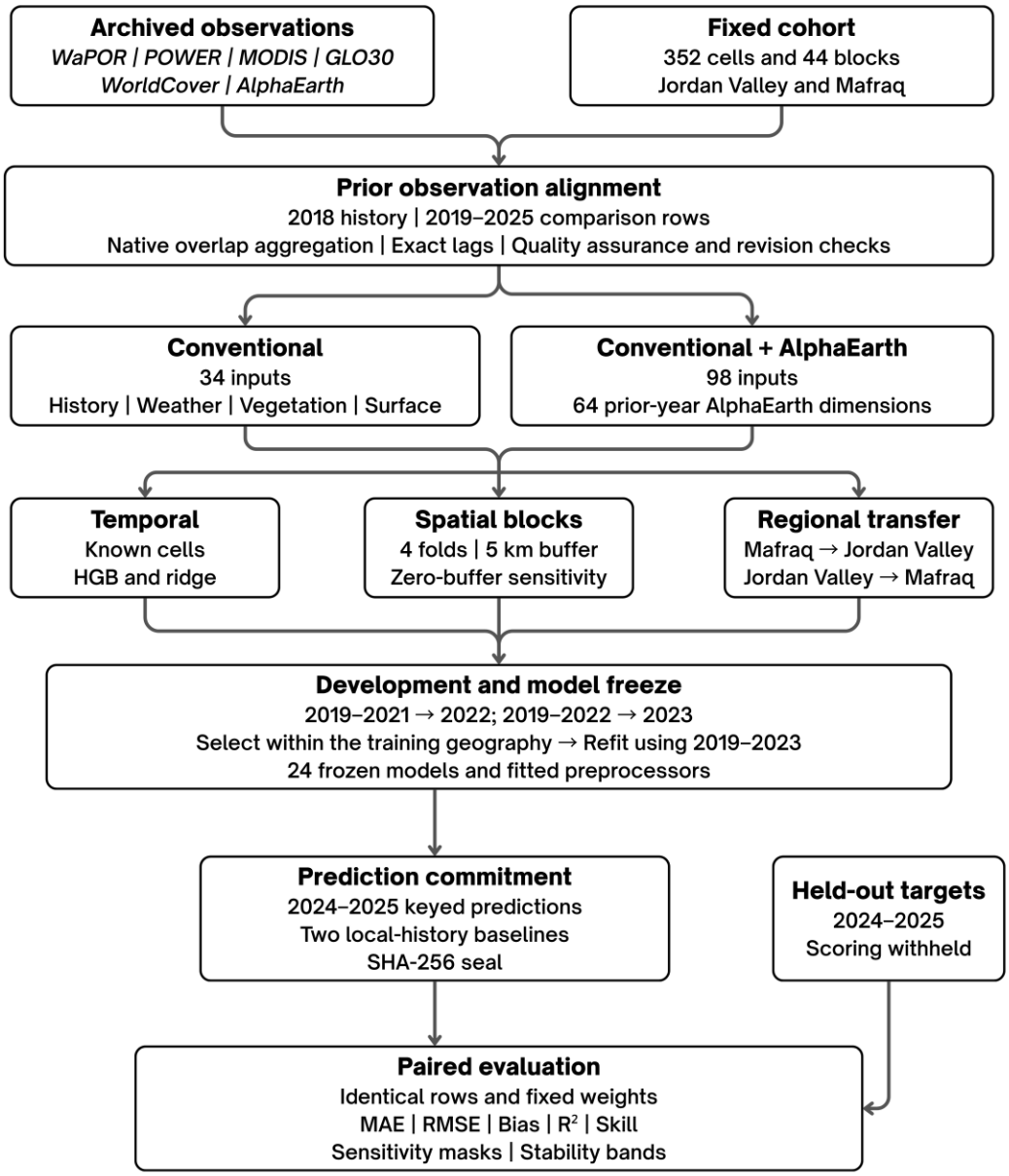


Figure 2. Observation Timing and Decision Locks in the Evaluation Workflow

Each scenario used its own permitted training geography. The 24 frozen models include eight zero-buffer sensitivity refits, which reused buffered settings. HGB = histogram gradient boosting; QA = quality assurance.

Data sources and cohort definition

The archived panel contained 33,792 cell-months for January 2018–December 2025 and 2,816 cell-year embedding records for 2017–2024. Original data were reused, with source identities, retrieval records, and SHA-256 digests retained. Public access routes and native spatial supports are identified in Table 2. WaPOR integers were converted once using a 0.1 multiplier to millimetres per calendar month; the original 1/1024° cell grid was preserved (Food and Agriculture Organization of the United Nations [FAO], 2026b).

The fixed cohort comprised 352 cells in 44 original 5-km blocks, with eight cells per block: 160 cells in the Jordan Valley and 192 in Mafraq. Inclusion required at least 70% WorldCover 2021 class-40 area and 95% valid mapped area, calculated by fractional pixel overlap. Class 40 primarily represents annual herbaceous cropland; irrigation and continuous cultivation cannot be inferred from this screen (European Space Agency [ESA], 2026). All original cells were retained. An 88-cell, seeded, single-AI review of 2021 annual Sentinel-2 composites provided plausibility screening only; neither exclusions nor predictors were derived from those judgments.

Table 2. Data Sources and Native Supports Used in Protocol v1

Product and support	Archived period and use	Public source and reproducibility identifier
WaPOR monthly AETI Nominal 100 m	2018–2025; target and local history. Native grid origin (−180°, 60°); step 1/1024°.	FAO (2026), WaPOR data portal; archived native integers, scale 0.1, nodata −9999.
Satellite Embedding V1 10 m; 64 dimensions	2017–2024; area means renormalized to unit length; previous year only.	Google (2026b); GOOGLE/SATELLITE_EMBEDDING/V1/ANNUAL. Google Cloud Storage: alphaearth_foundations.
ESA WorldCover 10 m	2021 v200 inclusion; 2020 v100 sensitivity screen. 2021 Sentinel-2 annual composites for	ESA (2026), WorldCover data access; class 40, fractional overlap; editions retained separately.

	screening.	
NASA POWER 0.5° × 0.625° meteorology; 1° solar	Daily UTC, 2017–2025; T2M, PRECTOTCORR, RH2M, WS2M, ALLSKY_SFC_SW_DWN.	NASA POWER Project (2026); daily point API. Raw responses, timestamps and hashes retained for four meteorology and two solar cells.
MOD13Q1.061 Terra 16 days; nominal 250 m	96 monthly cutoff slots, 2018–2025; preceding complete composite and strict QA.	Google (2026a); tile h21v05. Microsoft (n.d.) collection modis-13Q1-061; ORNL DAAC (2018), REST v1.
Copernicus DEM GLO30 Nominal 30 m DSM	Static surface elevation and Horn slope; principal acquisitions 2011–2015, older infill possible.	Copernicus Data Space Ecosystem (2026); public AWS GLO30 COG mirror; EGM2008 vertical reference.

Spatial support denotes the source footprint. The MODIS grid spacing is approximately 231.656 m in its sinusoidal projection.

Spatial harmonization and predictor timing

Native pixel–cell overlap areas were used for spatial aggregation. For AlphaEarth, the 64-dimensional annual vectors were averaged over each cell and renormalized, as expressed in Equation (1), where a_{ip} denotes overlap area and $v_{p\gamma}$ the native vector. All 64 dimensions were retained; the archived seam repair was preserved. Only the preceding calendar year was joined to each target year (Brown et al., 2025; Google, 2026b). Copernicus GLO30 supplied mean surface elevation and Horn slope, with $EGM2008$ heights, metric correction, and complete slope stencils; it was not treated as bare-earth terrain (Copernicus Data Space Ecosystem, 2026).

$$\overline{v}_{iy} = \frac{\sum_p a_{ip} v_{py}}{\sum_p a_{ip}}, z_{iy} = \frac{\overline{v}_{iy}}{\|\overline{v}_{iy}\|_2} \quad (1)$$

NASA POWER daily UTC records were summarized into monthly temperature, corrected precipitation, humidity, wind, and solar radiation. Meteorology was shared across four $0.5° \times 0.625°$ cells and radiation across two 1° cells, without field-scale downscaling (NASA

POWER Project, 2026). Precipitation was summed; multi-month temperature and other mean weather quantities were day-weighted. Exact-calendar AETI lags 1–12 and preceding 3-, 6-, and 12-month means were constructed. The conventional whitelist contained 34 inputs; augmentation added A00–A63, giving 98 inputs (Table S1). Identifiers, coordinates, region labels, and quality metadata were excluded.

$MOD13Q1.061$ normalized difference vegetation index (NDVI) was obtained through Microsoft Planetary Computer and ORNL DAAC (2018). The latest canonical Terra composite satisfying start plus 16 days ≤ target-month start was selected. Native values were multiplied by 0.0001. Every positive-area intersecting pixel had to satisfy the fixed range, reliability, cloud, aerosol, land, snow, shadow, and observation-date rules (Google, 2026a). Partial support remained missing. The September 2021 mismatch was resolved against the exact archived production revision before protocol locking; its April 2026 production date reinforces the retrospective interpretation.

Missing data and validation partitions

January–December 2018 supplied history only; 29,568 rows from 2019–2025 were eligible for comparison. Missing NDVI was replaced by one weighted median fitted exclusively within each training split, with the same missingness indicator included in both feature sets. No interpolation, stale carry-forward, or zero substitution was permitted. Unexpected missing required predictors, targets, or predictions stopped the corresponding release; targets were never imputed. Full-support NDVI and the pre-existing 264-cell two-edition WorldCover screen defined separate sensitivity masks applied to the primary predictions without refitting.

Development used two expanding origins, training on 2019–2021 and validating 2022, then training on 2019–2022 and validating 2023. After selection, each model was fitted once on its permitted 2019–2023 geography and held fixed throughout 2024–2025; the full temporal evaluation contained 8,448 rows. Earlier evaluation observations could populate later lag features, but could not update fitted components. No random row split, automatic early-stopping

split, or additional temporal embargo was used; overlapping historical windows were acknowledged.

Spatial blocking followed the prediction setting rather than assuming independent observations (Roberts et al., 2017). Within each region, blocks were sorted by northing then easting and divided into four rank groups. Each fold withheld 11 blocks and 88 cells. In $EPSG{:}32636$, block envelopes were enlarged to include cell footprints and padded by 1 m; whole training blocks within 5 km of a test envelope were excluded in every year. Tuning used only permitted training geography. Both regional-transfer directions were evaluated separately. Zero-additional-buffer sensitivity reused corresponding buffered settings without retuning (Table S2); buffers did not eliminate shared-weather dependence.

Models and training

Histogram gradient boosting (HGB) was the prespecified primary family; ridge regression provided a secondary temporal comparison (Friedman, 2001; Hoerl & Kennard, 1970). Persistence and seasonal-naive baselines used the preceding month and the same month one year earlier. HGB used squared-error loss, learning rate 0.05, 200 iterations, and minimum leaf size 50. Four candidates crossed 7 or 15 leaves with L2 penalties of 1 or 10. Ridge used an intercept, the SVD solver, and penalties of 1, 10, 100, or 1,000. Complete estimator settings are provided in Table S3.

Each feature set received the same search budget. Selection minimized the equally weighted mean of the two annual development MAEs; predefined ties favored simpler HGB trees, then stronger regularization. Ridge continuous inputs were standardized using weighted training means and population variances after imputation; the missingness indicator remained unscaled. HGB was unscaled. No feature selection, target transformation, or foundation-model retraining was performed. All predictions were clipped at zero without an upper cap, retaining raw values for auditing. The recorded workload comprised 128 development fits and 24 frozen refits.

Weighting and evaluation measures

Equation (2) assigns equal mass to active regions, blocks within region, months within block, and cells within block-month. R, B_r, M_{rb}, and C_{rbt} denote those counts. Training weights were rescaled to mean one and used for fitting, NDVI imputation, and ridge scaling. These weights describe the sampled cohort, not national water consumption. Spatial predictions were pooled across unique out-of-fold keys before scoring.

$$w_{\mathrm{rbtc}} = \frac{1}{RB_r M_{\mathrm{rb}} C_{\mathrm{rbt}}}, \sum_i w_i = 1 \tag{2}$$

With $e_i = \hat{y}_i - y_i$ and normalized weights, Equation (3) defines mean absolute error (MAE), root mean squared error (RMSE), and signed bias. Equation (4) defines weighted R^2, seasonal-naive skill, and the paired MAE contrast; negative ΔMAE favors augmentation. Zero denominators yielded undefined values. MAE was primary; all comparisons retained identical keys and weights. Region, year, calendar month, block, weather-grid, and sensitivity summaries were prescribed, with denominators and empty groups disclosed.

$$\mathrm{MAE} = \sum_i w_i |e_i|, \mathrm{RMSE} = \sqrt{\sum_i w_i {e_i}^2}, \mathrm{Bias} = \sum_i w_i e_i \tag{3}$$

$$R^2 = 1 - \frac{\sum_i w_i {e_i}^2}{\sum_i w_i (y_i - \overline{y})^2}, \overline{y} = \sum_i w_i y_i$$

$$\mathrm{Skill} = 1 - \frac{\mathrm{MAE}_{\mathrm{model}}}{\mathrm{MAE}_{\mathrm{seasonal\ naive}}}, \Delta\mathrm{MAE} = \mathrm{MAE}_{\mathrm{augmented}} - \mathrm{MAE}_{\mathrm{conventional}} \tag{4}$$

For temporal and pooled buffered-spatial HGB contrasts, 2,000 paired resamples used seed 20261016. Blocks were sampled with replacement within each region, and eight calendar quarters were independently resampled using a common quarter draw across regions and blocks. Each selected quarter retained its three months, and each block its eight cells. The 2.5th–97.5th percentiles formed descriptive stability bands conditional on frozen models and sampled sites and years, excluding training uncertainty; no population-level confidence or significance claim was intended.

Computing reproducibility and research ethics

The resumed pipeline ran locally on a MacBook Pro with an 11-core Apple M3 Pro, 18 GB memory, and macOS 15.6. Python 3.13.7 and scikit-learn 1.9.1 were used with one model job and at most two CPU threads. Cached downloads and fits were reused. Separate processes generated predictions and scored outcomes: all keyed predictions and frozen artifacts were hash-committed before scoring, with fitting disabled. This was application-level isolation, not cryptographic concealment of information recoverable from permitted lags.

Code, the locked protocol, spatial assignments, and compact processed inputs have been uploaded to the private JordanET GitHub repository (https://github.com/SamerAbaddi/JORDANET; commit $126410575e34$). Access requires the owner's permission. The file manifest records SHA-256 digests, and a restoration utility reconstructs the original working paths from the publication names. The supplied tables support computational reproduction; full raw-source reconstruction additionally requires the acquisition evidence archives identified in Table S4. Upstream data terms remain applicable.

Only public Earth-observation products were analyzed; no participants, personal records, or field interventions were involved. No institutional ethics approval is claimed. Source attribution and applicable data terms were retained, and satellite cells were not interpreted as verified farms or ownership boundaries. AI-assisted imagery screening was disclosed and was not presented as independent human validation.

Results

Under protocol $v1$, prior-year AlphaEarth embeddings did not reduce WaPOR AETI error in the primary temporal evaluation. The conventional histogram gradient boosting (HGB) model achieved the lowest mean absolute error (MAE) among the six temporal methods. Regional transfer showed a different pattern: augmentation improved both transferred HGB models, although transfer into Mafraq remained less accurate than the local-history baselines.

The following comparisons separate these findings across development, temporal evaluation, spatial holdouts, and the prescribed sensitivity analyses.

Evaluation coverage and development

All 352 study cells and 44 original blocks were retained. The 2024–2025 evaluation comprised 8,448 unique cell-months: 3,840 in the Jordan Valley and 4,608 in Mafraq. No prediction failed, no evaluation row was dropped, and no target was imputed. All 152,064 method–scenario prediction records were sealed before outcome scoring; their larger count reflects repeated observations across methods and settings, rather than additional independent samples.

Full-support vegetation observations were available for 7,810 evaluation rows; the remaining 638 rows were retained using the frozen training-based NDVI imputation and missingness flag. The two-edition cropland sensitivity included 264 cells and 6,336 rows. These restrictions changed the population being summarized but did not change fitted models or predictions.

On the 2022–2023 development observations, HGB MAE decreased from 5.7546 to 5.7278 mm/month with augmentation, a difference of -0.0268 mm/month (-0.47%). Ridge MAE instead increased from 6.3300 to 6.4217 mm/month. These values were used in model selection and therefore represent development diagnostics. The temporal and restricted-geography development comparisons are reported in Table S5.

Primary temporal comparison

The small HGB development advantage was not retained in final evaluation. Temporal MAE increased from 6.4298 to 6.5850 mm/month after AlphaEarth was added, giving ΔMAE = +0.1552 mm/month (+2.41%). Root mean squared error (RMSE) increased from 10.0596 to 10.2062 mm/month, while weighted R^2 decreased from 0.8451 to 0.8406. Signed bias was slightly lower, changing from +0.2837 to +0.2664 mm/month; this reduction in average signed error did not correspond to lower absolute error (Table 3).

Table 3. Frozen Temporal Models and Local History Baselines on the Final Evaluation

Method	MAE	RMSE	Bias	R^2	Skill
HGB conventional	6.430	10.060	0.284	0.845	0.323
HGB + AlphaEarth	6.585	10.206	0.266	0.841	0.307
Ridge conventional	6.596	10.249	0.473	0.839	0.306
Ridge + AlphaEarth	6.663	10.277	0.623	0.838	0.299
Persistence	8.684	14.122	0.158	0.695	0.086
Seasonal naive	9.500	15.906	1.836	0.613	0.000

MAE, RMSE, and bias are in mm/month; R^2 and seasonal-naive skill are dimensionless. Metrics use the protocol's region–block–month–cell weighting. Bias is prediction minus reference.

Conventional HGB reduced MAE by 32.32% relative to seasonal naive and by 25.96% relative to persistence. The augmented HGB model also outperformed both baselines, but its seasonal-naive skill was lower, at 0.3069 compared with 0.3232. Thus, the main predictive gain over simple local history was already present in the conventional feature set. Figure 3 places this temporal result alongside the spatial and regional-transfer comparisons.

Ridge gave the same direction of incremental change: MAE increased by 0.0665 mm/month and RMSE by 0.0281 mm/month. Before the prescribed zero clipping, 318 conventional-ridge and 296 augmented-ridge predictions were negative; all occurred in Mafraq. Neither temporal HGB model produced a negative raw prediction. Table 3 reports the clipped predictions used for every scored comparison, and Figure S1 shows the regional prediction distributions for HGB.

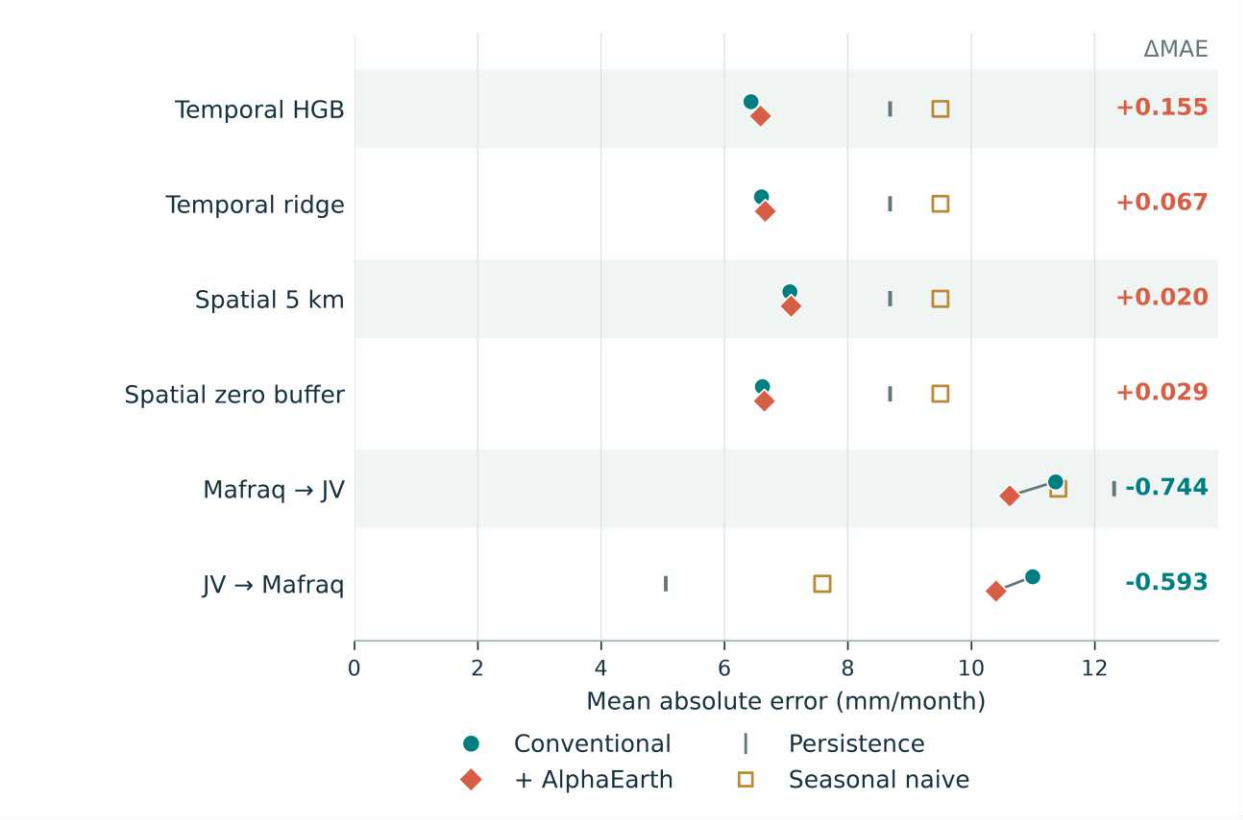


Figure 3. Incremental AlphaEarth Performance Across Temporal and Spatial Evaluation Settings

ΔMAE is augmented minus conventional; negative values favor augmentation.

Spatial holdouts and regional transfer

With the 5-km buffer, pooled HGB MAE changed from 7.0588 to 7.0787 mm/month (ΔMAE = +0.0199), whereas RMSE decreased slightly by 0.0099 mm/month. The mean absolute error contrast was positive in three folds and negative in one: +0.0124, +0.0811, +0.0575, and -0.0713 mm/month for folds 1–4, respectively. Fold 3 combined higher MAE with lower RMSE, showing that the two error measures did not move uniformly.

Removing the additional buffer reduced absolute MAE for both feature sets, but the augmented model again had slightly higher pooled MAE: 6.6440 versus 6.6151 mm/month. The incremental fold differences were +0.0453, +0.1591, -0.0925, and +0.0035 mm/month. This sensitivity changes both proximity to test locations and the amount of training data. It therefore does not isolate a causal buffer effect. The buffered training populations ranged from 7 to 23

blocks, whereas each zero-buffer fold retained 33 blocks; Figure S2 displays the corresponding buffered geography.

Table 4. Paired MAE Comparisons and Baseline Context Across Evaluation Settings

Setting	C	A	ΔMAE	P	S
Temporal HGB	6.430	6.585	0.155	8.684	9.500
Temporal ridge	6.596	6.663	0.067	8.684	9.500
Buffered spatial	7.059	7.079	0.020	8.684	9.500
Zero buffer	6.615	6.644	0.029	8.684	9.500
Mafraq → JV	11.365	10.620	-0.744	12.317	11.412
JV → Mafraq	10.994	10.400	-0.593	5.050	7.588

All entries are in mm/month. C = conventional; A = conventional + AlphaEarth; P = persistence; S = seasonal naive. JV = Jordan Valley. Negative ΔMAE favors augmentation.

When training in Mafraq and evaluating the Jordan Valley, augmentation reduced MAE from 11.3645 to 10.6203 mm/month, an improvement of 0.7442 mm/month (6.55%). RMSE decreased by 1.1283 mm/month, and negative bias was reduced from -4.2986 to -3.6319 mm/month. The augmented MAE was lower than both persistence (12.3170) and seasonal naive (11.4123), although substantial negative bias remained.

In the reverse direction, MAE decreased from 10.9937 to 10.4002 mm/month (5.40%), with RMSE decreasing by 0.7863 mm/month. The improvement was insufficient to match persistence at 5.0503 or seasonal naive at 7.5883 mm/month. Positive bias remained pronounced at +8.5728 mm/month, and R^2 was only 0.0838 after augmentation. The regional-transfer benefit was therefore relative to the transferred conventional model; it did not establish competitive performance in both destinations (Table 4).

Regional and seasonal error patterns

The unfavorable primary contrast occurred in both sampling areas and both evaluation years. Temporal HGB MAE increased from 8.3945 to 8.5695 mm/month in the Jordan Valley and from 4.4650 to 4.6006 mm/month in Mafraq. Annual ΔMAE was +0.0959 mm/month in

2024 and +0.2146 mm/month in 2025. The higher Jordan Valley error level was shared by both feature sets, while augmentation introduced a smaller absolute deterioration in Mafraq (Table S6). The regional mean trajectories were closely aligned between the two HGB models, yet their cell-level absolute-error differences changed sign between months (Figure 4).

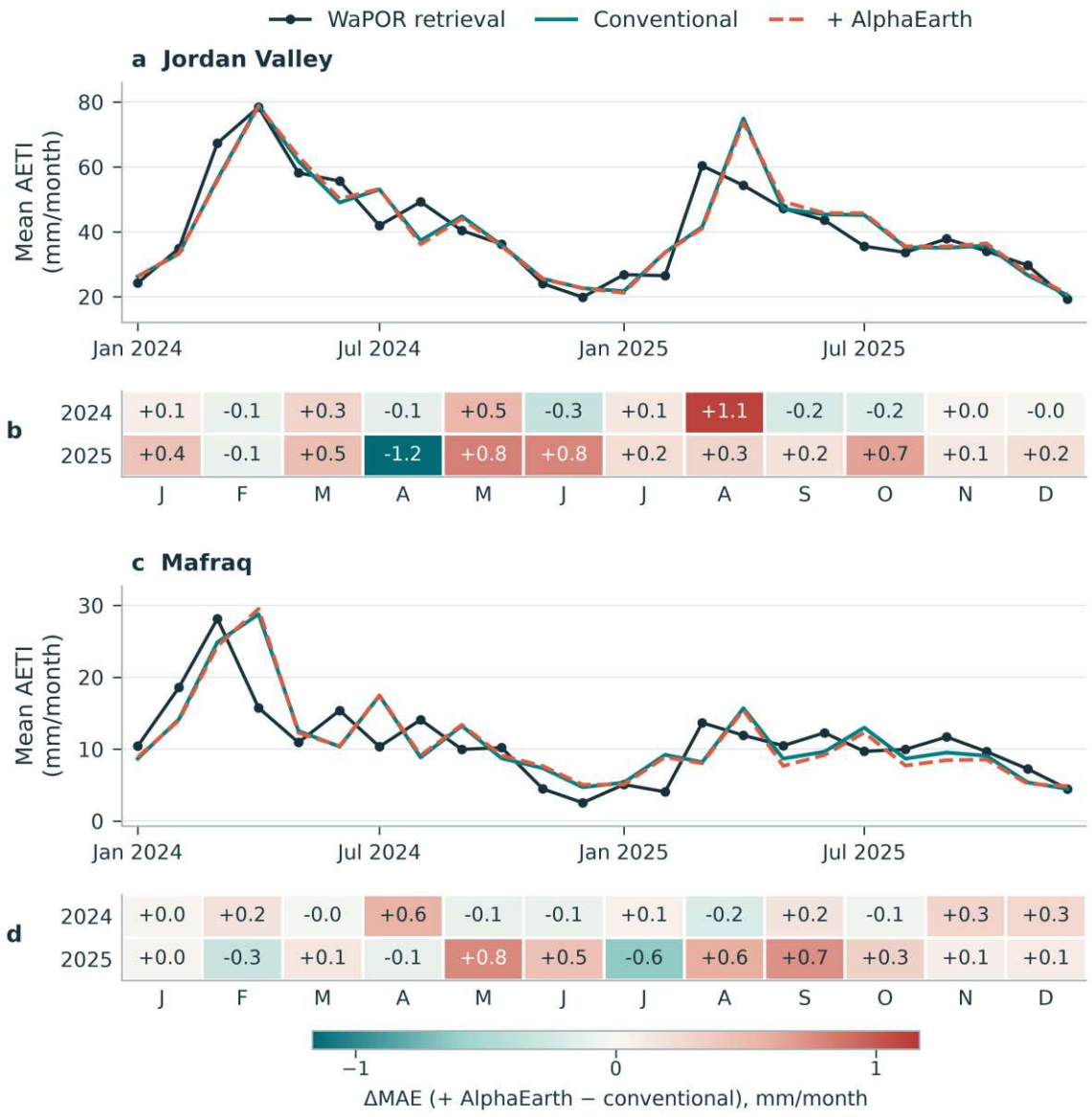


Figure 4. Seasonal AETI Trajectories and Monthly Incremental Error in the Two Regions

Panels a and c show regional monthly means; panels b and d show monthly ΔMAE for the temporal HGB pair.

Visible differences between predictions and WaPOR retrievals remained, including the timing of the early-2024 Mafraq peak. Agreement in regional averages did not imply uniformly accurate cell predictions.

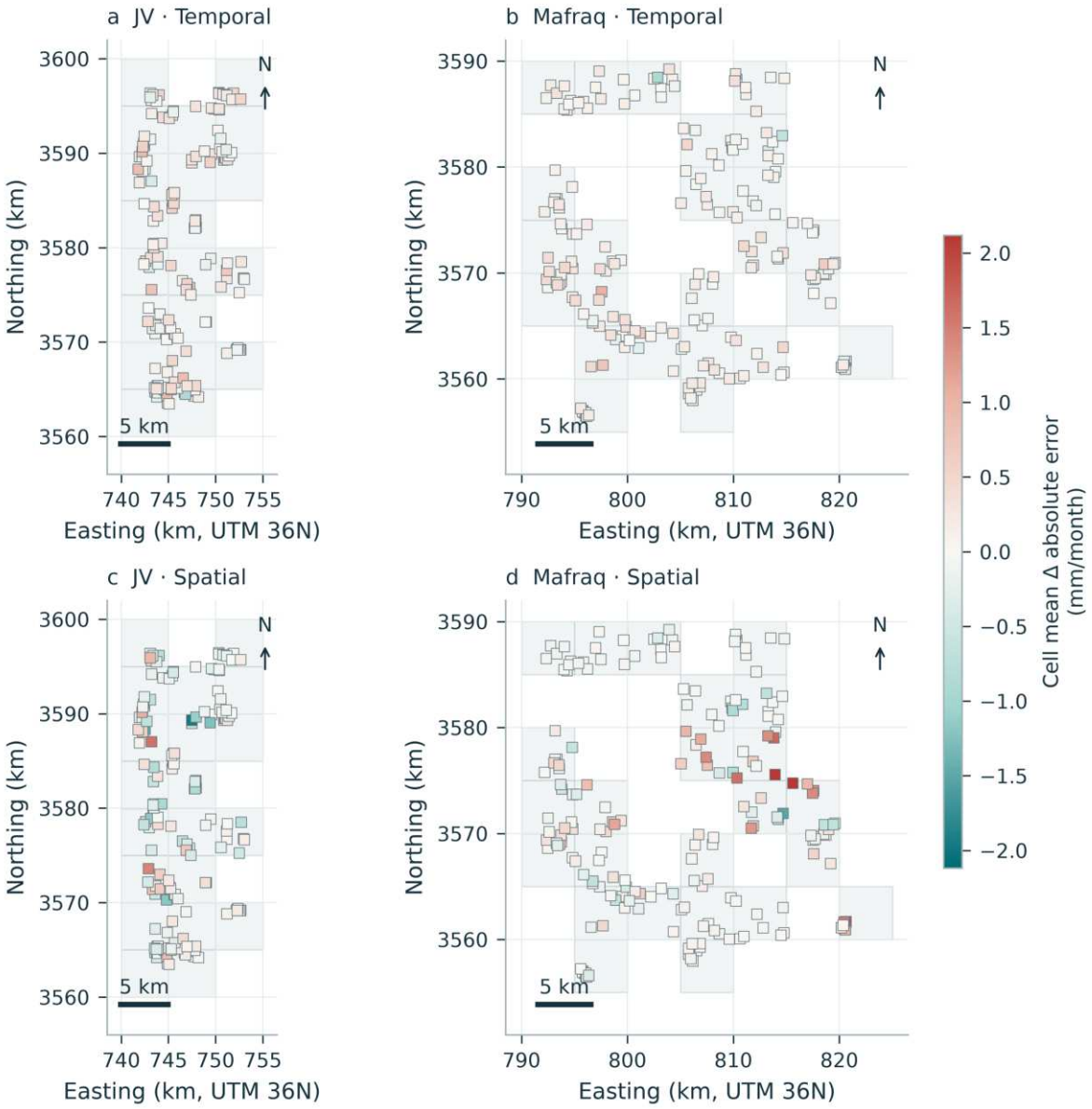


Figure 5. Sampled Cell Differences in Temporal and Buffered Spatial Prediction Error

Each symbol represents its cell's mean absolute-error difference over 2024–2025, augmented minus conventional HGB. Coordinates are UTM zone 36N.

The heatmaps therefore display monthly differences in individual-cell absolute errors separately from the regional mean curves. Buffered spatial differences also varied by region. The augmented MAE was lower by 0.0889 mm/month in the Jordan Valley and higher by 0.1287 mm/month in Mafraq. The pooled contrast was +0.0400 in 2024 and -0.0002 mm/month in 2025. Cell summaries show spatially intermixed improvements and deteriorations, with a wider visual range in the buffered setting than in the temporal setting (Figure 5). These maps describe sampled cells; they do not extend the evaluation to unsampled land.

Sensitivity and conditional stability

Restricting evaluation to full-support NDVI rows did not reverse the temporal result: HGB ΔMAE remained +0.1490 mm/month. The two-edition cropland screen yielded +0.1710 mm/month. Ridge remained unfavorable under both restrictions. Pooled buffered and zero-buffer HGB contrasts also remained positive for MAE, while both regional-transfer directions retained negative contrasts. The complete sensitivity comparisons and their changing denominators are shown in Table S7.

For the primary HGB contrast, the archived block-by-quarter resamples gave a 2.5th–97.5th percentile stability band of +0.0819 to +0.2322 mm/month around the observed +0.1552. The corresponding buffered-spatial band extended from -0.1903 to +0.2032 mm/month around +0.0199 (Figure 6). Each distribution contains 2,000 paired resamples of fixed predictions. These are descriptive bands conditional on the frozen models and sampled sites and years; they omit model-fitting and selection uncertainty and are not population confidence intervals or significance tests.

The archived evaluation audit checked 2,422 defined metric rows, 642 paired contrasts, and all 4,000 stability draws. Eight metric rows were explicitly undefined because the requested region was absent from a directional transfer test. All performance values quantify prediction of archived WaPOR AETI retrievals; independent field evapotranspiration was not observed.

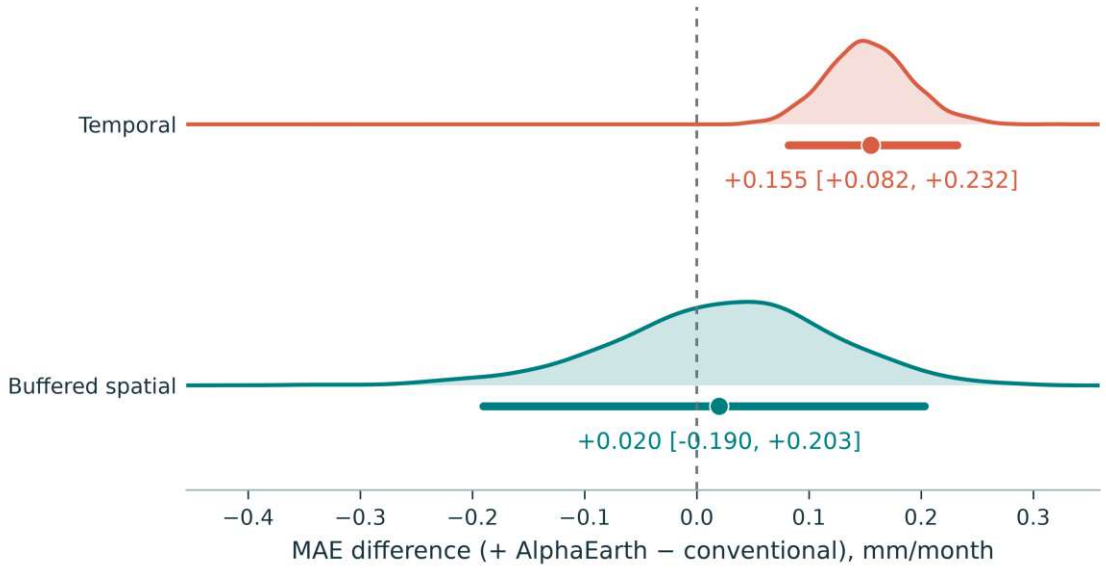


Figure 6. Conditional Stability of the Temporal and Buffered Spatial MAE Contrasts

Note. Dots mark the observed paired differences; horizontal intervals show the stored 2.5th–97.5th percentiles.

Discussion

AlphaEarth's contribution depended on where prediction was required and which comparator was used. In the primary temporal evaluation of WaPOR actual evapotranspiration and interception (AETI), adding prior-year embeddings increased histogram gradient boosting (HGB) mean absolute error (MAE) from 6.4298 to 6.5850 mm/month. Conventional HGB already reduced MAE by 25.96% against persistence and 32.32% against seasonal naive. These gains belong to the conventional prediction system. The embeddings improved both regional-transfer models, but their practical value differed sharply between destinations.

The temporal deterioration was modest but consistent across both regions, both evaluation years, and the secondary ridge comparison. A small development advantage did not survive the held-out period. One plausible explanation is overlap between information encoded by annual embeddings and the seasonality, antecedent AETI, weather, vegetation, and terrain already supplied to the conventional models. Previous-year descriptors cannot directly observe

new management changes during the target year. Additional dimensions may have complicated estimation within the limited training geography.

This result does not contradict AlphaEarth's original mapping benchmarks. Brown et al. (2025) evaluated monthly OpenET proxy prediction using observation support extending through the target month. Here, the target was next-month WaPOR AETI, and embeddings were restricted to the preceding year. The target product, information cutoff, and comparator were therefore different. Brown et al.'s benchmark already established an ET-related application; the contribution here is the controlled comparison against a history-rich predictor set.

A closer conceptual parallel is provided by Liu et al. (2026), who found similar temporal rainfall–runoff performance using AlphaEarth and conventional catchment descriptors, with stronger gains from their combination in ungauged-basin prediction. Both studies suggest that representation value changes with the generalization task. However, the Jordan experiments retained prior local AETI at held-out locations, whereas ungauged-basin prediction excludes observed discharge histories from the target basins.

Transfer into the Jordan Valley benefited from augmentation: MAE fell by 6.55%, and the augmented model outperformed both local-history baselines. In the reverse direction, the 5.40% reduction left MAE at 10.4002 mm/month, compared with 5.0503 for persistence, alongside substantial positive bias. Thus, improvement relative to the transferred conventional model did not ensure a competitive forecast in Mafraq. Siddiqui et al. (2026) similarly found that AlphaEarth's advantage over engineered features in irrigation classification varied between data-rich Uganda and data-sparse Zambia. Their classification results support context-specific assessment, without establishing the mechanism behind Jordan's directional asymmetry.

The buffered spatial results provide little evidence of an average MAE advantage. The pooled increase of 0.0199 mm/month coexisted with a slight decrease in root mean squared error and mixed fold-level changes. Its conditional stability band crossed zero; equivalence was not established. By comparison, the temporal band remained above zero, indicating a stable unfavorable contrast under the prescribed resampling, without providing a population

significance test. Removing the additional buffer lowered absolute errors while increasing the available training population. Consistent with Roberts et al. (2017), these differences must be interpreted in relation to the prediction setting; proximity and training-size effects were not isolated experimentally.

Regional and monthly diagnostics qualify the pooled findings. Similar regional mean trajectories concealed cell-level improvements and deteriorations, while the larger Jordan Valley errors were shared by both feature sets. Crop calendars, management, and retrieval behavior could contribute, but no farm observations were available to attribute the differences. Restricting evaluation to full-support vegetation records or the two-edition cropland screen preserved the direction of the primary contrast. The deterioration was therefore not confined to rows requiring vegetation imputation, although these checks cannot establish that all covariate or land-cover errors were removed.

All reported accuracy concerns reproduction of WaPOR retrievals. In the North Jordan Valley, Al-Bakri et al. (2022) showed that WaPOR overestimation relative to calibrated energy-balance estimates affected agricultural water-accounting outputs. Those findings motivate independent validation; they do not quantify bias in the present archive. Volk et al. (2024), by contrast, evaluated OpenET against in situ observations across the United States. Errors from such field validation cannot be ranked against the smaller errors obtained here when predicting a satellite product. A model could reproduce its reference closely while retaining the reference's physical inaccuracies.

Implications

For representation learning, the findings support treating embedding value as conditional on the information already available and the geography of deployment. A useful representation need not improve an already informative local model. This distinction narrows the theoretical contribution to an empirical boundary on added predictive value; neither a universal advantage nor an intrinsic failure of foundation embeddings is supported. Future benchmarks should make

this conditional question explicit by separating representation replacement, augmentation, and prediction without local target history.

Methodologically, equal comparison rows, training-only preprocessing, fixed selection budgets, and sealed predictions make unfavorable findings interpretable. Engineering value should likewise be assessed through both prediction quality and the monitoring workflow. In a recent power-load study, Abaddi (2026a) coupled a seasonal statistical model with an AI residual learner and a decision-support dashboard; hybrid accuracy remained comparable to the statistical baseline. That cross-domain example reinforces the need to distinguish the value of a complete information system from the incremental benefit of its AI component. For Jordan, deployment assessment should include data latency and the cost of forecast errors.

For water policy, a reasonable next application is a monitored pilot supporting review of agricultural water-use patterns, subject to independent validation. The current predictions do not identify irrigation demand, abstraction, or achievable water savings. Grafton et al. (2018) showed why efficiency interventions require basin accounting and attention to behavioral responses before savings can be inferred. Consequently, forecasts should be combined with measured deliveries and consumption, explicit uncertainty, and local expertise. Equal-region evaluation weights also provide no basis for estimating national water volumes or allocating farm-level entitlements.

Limitations and future directions

Generalization is limited by 352 cropland-screened cells in 44 blocks across two sampling areas and a two-year final evaluation. Repeated cell-months are not independent geographic replicates. Some buffered fits retained very few blocks, and shared coarse weather grids preserved dependence despite spatial separation. Conditional resampling omitted model-fitting and selection uncertainty, and no stability intervals were prescribed for regional transfer.

The observation-time restrictions did not establish historical product-release availability or foundation-model pretraining cutoffs. The corrected vegetation revision was produced after

the observation period, so this remains a retrospective experiment rather than a demonstrated operational forecast. In PHASE-ED, shorter recorded recovery-period stays were observed before a change in timestamp precision, highlighting the need to examine temporal trends alongside changes in recording practice (Abaddi, 2026b). For satellite-based forecasting, this supports auditing product revisions when interpreting temporal performance, without assuming that revisions explain the observed model differences.

The compact search covered HGB and a secondary temporal ridge comparison without feature selection or foundation-model adaptation. Each proposal now requires a new evaluation set because the current outcomes have been examined. Advancement should depend on an independently confirmed, decision-relevant improvement over local-history baselines, followed by evidence that using the forecasts improves water-management decisions.

Conclusion

Prior-year AlphaEarth embeddings did not improve next-month WaPOR AETI prediction in the primary temporal evaluation. Conventional histogram gradient boosting provided the lowest temporal MAE among the tested methods, while augmentation increased error in both boosting and ridge regression. The pooled buffered spatial comparison also provided little evidence of an average MAE benefit. Within this pipeline, annual embeddings therefore offered no general advantage over conventional predictors that already included local AETI history.

Regional transfer revealed a narrower benefit. Embeddings improved the transferred model in both directions, but only transfer into the Jordan Valley surpassed both local-history baselines. The reverse transfer remained substantially less accurate than persistence. This asymmetry makes the choice of comparator central to interpretation: improvement over a transferred model does not, by itself, establish a useful prediction system at the destination.

The contribution is a controlled assessment of added embedding value across temporal and spatial settings, with equal comparison rows, frozen models and predictions sealed before scoring. Its scope is the sampled cohort and the WaPOR reference product. Progress toward

operational use requires independent evapotranspiration measurements, documented release-time availability and replication across additional agricultural settings. Any subsequent model changes should be evaluated on new outcomes because the present evaluation has been examined.

Declarations

Conflict of interest

The author declares no financial or non-financial competing interests.

Funding

The study received no external funding.

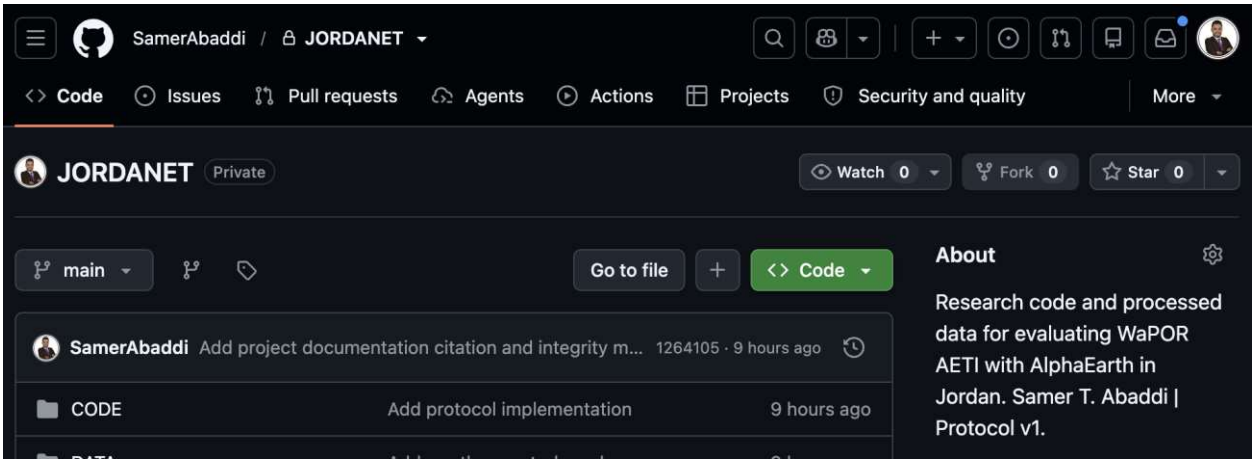
Data availability

Public source products comprise FAO WaPOR, Google Satellite Embedding V1, ESA WorldCover, NASA POWER, MOD13Q1.061 and Copernicus DEM GLO30. Source access routes, versions and processing rules are documented in the Methodology and JordanET.

Four videos were shared with this submission. Video 1 presents the private JORDANET GitHub repository, showing the organization of the study code, processed data, protocol, environment files and workflow figures. Video 2 demonstrates navigation of the official FAO WaPOR portal, including evapotranspiration and interception products, geographic searches and temporal controls, supporting the manuscript's data-source description. Video 3 shows the official AlphaEarth dataset entry in the Google Earth Engine Data Catalog, highlighting dataset availability, image properties and citation information. Video 4 opens the JORDANET results notebook in Google Colab and displays the archived evaluation table and comparison plot for conventional models.

Code availability

The study code, protocol, software specifications and reproduction instructions are deposited in the private JordanET GitHub repository at https://github.com/SamerAbaddi/JORDANET. The verified repository version is commit 126410575e34019b4b5142e522fa1a5239df097f. Exceptional access to the GitHub repository can be granted at the Editor-in-Chief's request. Post-acceptance, the repository is intended to be public and linked to the article DOI's.



Ethics approval

This study analyzed secondary environmental data and satellite imagery. No human participants, identifiable personal information, animals or field interventions were involved. No institutional ethics approval or formal exemption is claimed.

Consent to participate and publication

Not applicable. No participant records or identifiable personal material were used.

Clinical trial registration

Not applicable. This was an environmental prediction study and involved no clinical intervention.

AI assistance

AI was used for proofreading, improving the flow and correcting typos. Responsibility for the scientific content and final manuscript rests with the author.

References

Abaddi, S. (2026a). The optimization and load forecasting of power distribution units in data halls. Applied Operations and Analytics, 2(1), Article 2649444. https://doi.org/10.1080/29966892.2026.2649444

Abaddi, S. (2026b). PHASE-ED: Recorded emergency department stay and temporal prediction across pandemic phases at Jordanian public hospitals [Preprint]. Research Square. https://doi.org/10.21203/rs.3.rs-11366553/v1

Abaddi, S. T. (2025). Mush-Room for innovation: Internet of Things, machine learning, and generative pre-trained transformers integration. Cureus Journal of Computer Science, 2, Article es44389-025-03223-1. https://doi.org/10.7759/s44389-025-03223-1

Al-Bakri, J. T., D'Urso, G., Batchelor, C., Abukhalaf, M., Alobeiaat, A., Al-Khreisat, A., & Vallee, D. (2022). Remote sensing-based agricultural water accounting for the North Jordan Valley. Water, 14(8), Article 1198. https://doi.org/10.3390/w14081198

Blatchford, M. L., Mannaerts, C. M., Njuki, S. M., 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

Brown, C. F., Kazmierski, M. R., Pasquarella, V. J., Rucklidge, W. J., Samsikova, M., Zhang, C., Shelhamer, E., Lahera, E., Wiles, O., Ilyushchenko, S., Gorelick, N., Zhang, L. L., Alj, S., Schechter, E., Askay, S., Guinan, O., Moore, R., Boukouvalas, A., & Kohli, P. (2025). AlphaEarth Foundations: An embedding field model for accurate and efficient global mapping from sparse label data [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2507.22291

Chattha, M. K., Youssef, H., Ftomova, O., Maseeh, A. N., Wang, X., Bogetić, Ž., Naeher, D., Borja-Vega, C., & Ghosheh, A. (2024). Gulf economic update, December 2024: Navigating the water challenge in the GCC—Paths to sustainable solutions. World Bank.

https://documents.worldbank.org/en/publication/documents-reports/documentdetail/099002211272487299

Copernicus Data Space Ecosystem. (2026). Copernicus DEM—Global and European digital elevation model. https://dataspace.copernicus.eu/explore-data/data-collections/copernicus-contributing-missions/collections-description/COP-DEM

de Waal, D., Khemani, S., Barone, A., & Borgomeo, E. (2023). The economics of water scarcity in the Middle East and North Africa: Institutional solutions. World Bank. https://doi.org/10.1596/978-1-4648-1739-7

European Space Agency. (2026). Data access. ESA WorldCover. https://esa-worldcover.org/en/data-access

Food and Agriculture Organization of the United Nations. (2026a). WaPOR, remote sensing for water productivity. https://www.fao.org/land-water/resources/tools/databases/wapor/en

Food and Agriculture Organization of the United Nations. (2026b). WaPOR data. https://www.fao.org/in-action/remote-sensing-for-water-productivity/wapor-data/en

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. The Annals of Statistics, 29(5), 1189–1232. https://doi.org/10.1214/aos/1013203451

Google. (2026a). MOD13Q1.061 Terra vegetation indices 16-day global 250m. Earth Engine Data Catalog. https://developers.google.com/earth-engine/datasets/catalog/MODIS_061_MOD13Q1

Google. (2026b). Satellite Embedding V1. Earth Engine Data Catalog. https://developers.google.com/earth-engine/datasets/catalog/GOOGLE_SATELLITE_EMBEDDING_V1_ANNUAL

Grafton, R. Q., Williams, J., Perry, C. J., Molle, F., Ringler, C., Steduto, P., Udall, B., Wheeler, S. A., Wang, Y., Garrick, D., & Allen, R. G. (2018). The paradox of irrigation efficiency. Science, 361(6404), 748–750. https://doi.org/10.1126/science.aat9314

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 12(1), 55–67. https://doi.org/10.1080/00401706.1970.10488634

Liu, J., Shen, C., Yang, Y., Ji, H., Lawson, K., & Kraabel, N. (2026). Geospatial foundation embeddings as transferable catchment descriptors. Geophysical Research Letters, 53(16), Article e2026GL122814. https://doi.org/10.1029/2026GL122814

Microsoft. (2026). MODIS vegetation indices 16-day (250m). Planetary Computer. https://planetarycomputer.microsoft.com/dataset/modis-13Q1-061

Ministry of Water and Irrigation. (2023). National water strategy 2023–2040. https://mwi.gov.jo/EBV4.0/Root_Storage/AR/EB_List_Page/national_water_strategy_2023-2040.pdf

Mughees, M. A., Montefoschi, G., Chen, Z., & Brovelli, M. A. (2026). From foundation embeddings to cropland maps: Label efficiency, temporal transferability and independent human validation [Preprint]. arXiv. https://arxiv.org/abs/2609.17138

NASA POWER Project. (2026). Data sources. https://power.larc.nasa.gov/docs/methodology/data/sources/

Natural Earth. (2026). Admin 0—Countries (Version 5.1.1) [Data set]. https://www.naturalearthdata.com/downloads/50m-cultural-vectors/50m-admin-0-countries-2/

ORNL DAAC. (2018). Terrestrial Ecology Subsetting & Visualization Services (TESViS) RESTful Web Service. https://doi.org/10.3334/ORNLDAAC/1600

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F., & Dormann, C. F. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography, 40(8), 913–929. https://doi.org/10.1111/ecog.02881

scikit-learn developers. (2026). Supervised learning. scikit-learn 1.9.1 documentation. https://scikit-learn.org/stable/supervised_learning.html

Siddiqui, H., Small, C., & Modi, V. (2026). From foundational models to interpretability: A comparative study of embeddings and feature engineering for smallholder irrigation

detection in Sub-Saharan Africa. Frontiers in Remote Sensing, 7, Article 1895172. https://doi.org/10.3389/frsen.2026.1895172

United Nations Educational, Scientific and Cultural Organization. (2024, April 24). Food security, a key driver of peace and prosperity. https://www.unesco.org/reports/wwdr/en/2024/agriculture

Volk, J. M., Huntington, J. L., Melton, F. S., Allen, R., Anderson, M., Fisher, J. B., Kilic, A., Ruhoff, A., Senay, G. B., Minor, B., Morton, C., Ott, T., Johnson, L., Comini de Andrade, B., Carrara, W., Doherty, C. T., Dunkerly, C., Friedrichs, M., Guzman, A., . . . Yang, Y. (2024). Assessing the accuracy of OpenET satellite-based evapotranspiration data to support water resource and land management applications. Nature Water, 2(2), 193–205. https://doi.org/10.1038/s44221-023-00181-7

Supplementary materials

Table S1. Locked Predictor Groups and Observation Timing

Group and count	Exact fields or indexed field families
Calendar · 3	month_sin; month_cos; target_days
Surface · 2	elevation_mean_m_egm2008; slope_mean_degrees
AETI history · 15	aeti_lag01_mm through aeti_lag12_mm; aeti_prior03_mean_mm; aeti_prior06_mean_mm; aeti_prior12_mean_mm
Temperature and rain · 6	temperature_lag01_C; precipitation_lag01_mm; temperature_prior03_dayweighted_mean_C; precipitation_prior03_sum_mm; temperature_prior12_dayweighted_mean_C; precipitation_prior12_sum_mm
Additional weather · 6	lag1_RH2M_percent; lag1_WS2M_m_s; lag1_solar_MJ_m2_day; prior3_RH2M_percent; prior3_WS2M_m_s; prior3_solar_MJ_m2_day
Vegetation · 2	ndvi_last_complete_16d_full_support; ndvi_missing_flag
Embeddings · 64	A00 through A63

Note. The conventional set contains 34 inputs and the augmented set 98.

Table S2. Spatial Partitions and Final Refit Sample Sizes

Scenario	Train blocks JV / Mafraq	Buffer blocks	Train cells Rows	Test cells Rows
Temporal	20 / 24	0	352 21,120	352 8,448
5-km fold 1	10 / 9	14	152 9,120	88 2,112
5-km fold 2	5 / 6	22	88 5,280	88 2,112
5-km fold 3	5 / 2	26	56 3,360	88 2,112
5-km fold 4	10 / 13	10	184 11,040	88 2,112
Zero-buffer each fold	15 / 18	0	264 15,840	88 2,112

Mafraq → JV	0 / 24	0	192 11,520	160 3,840
JV → Mafraq	20 / 0	0	160 9,600	192 4,608

Table S3. Selected Settings Retained in the Frozen Models

Scenario	Family	Conventional	Augmented
T_PRIMARY	HGB	Leaves 15; L2 1	Leaves 15; L2 1
T_PRIMARY	Ridge	α = 1000	α = 1000
B5000_F1	HGB	Leaves 7; L2 1	Leaves 7; L2 10
B5000_F2	HGB	Leaves 15; L2 1	Leaves 7; L2 10
B5000_F3	HGB	Leaves 15; L2 10	Leaves 7; L2 10
B5000_F4	HGB	Leaves 7; L2 10	Leaves 7; L2 10
B0_F1	HGB	Leaves 7; L2 1	Leaves 7; L2 10
B0_F2	HGB	Leaves 15; L2 1	Leaves 7; L2 10
B0_F3	HGB	Leaves 15; L2 10	Leaves 7; L2 10
B0_F4	HGB	Leaves 7; L2 10	Leaves 7; L2 10
LRO_JV	HGB	Leaves 15; L2 10	Leaves 7; L2 1
LRO_Mafraq	HGB	Leaves 7; L2 1	Leaves 15; L2 1

Table S4. Files Governing Computational Reproduction

Layer	Path and function
Protocol and comparison keys	outputs/Jordan_Evaluation_Protocol_v1/: PROTOCOL.json, PROTOCOL_LOCK.json, INPUT_LOCK.json, FEATURE_WHITELISTS.json, comparison_rows.csv.gz,

	block_folds.csv, cell_assignments.csv, spatial_roles.csv, run_schedule.csv.
Locked modeling inputs	work/restored/core/panel/baseline_panel.csv.gz and alphaearth_annual.csv.gz; work/restored/checkpoint/work/weather-controls/tables/weather_controls_lagged.csv.gz; work/resume/vegetation/vegetation_candidate_sidecar.csv.gz.
Execution and integrity	work/resume/training/train_protocol_v1.py; work/resume/final_evaluation/evaluate_v1.py. Input and protocol hashes checked before fitting; outcome-column projection tested; model fits disabled during final prediction; prediction hashes verified before scoring.
Model and prediction records	RUNTIME.json; requirements-training.txt; SELECTED_MODELS.json; frozen model records; PREDICTION_SEAL.json; PRE_SCORE_CHECKS.json. A seal commits to all 11 scenario prediction files before any final score is released.

Table S5. Temporal Development Diagnostics Used in Model Selection and Restricted Geography Development MAE and Local History Baselines

Method	**MAE**	**RMSE**	**Bias**	R^2	**Skill**
HGB conventional	5.755	9.610	-0.266	0.861	0.277
HGB + AlphaEarth	5.728	9.528	-0.299	0.864	0.280
Ridge conventional	6.330	10.509	-1.174	0.834	0.204
Ridge + AlphaEarth	6.422	10.584	-0.889	0.832	0.193
Persistence	8.489	14.231	-0.198	0.696	-0.067
Seasonal naive	7.954	14.029	0.201	0.705	0.000
Training setting	**Rows**	**C**	**A**	**P**	**S**
Buffered fold 1	3,648	5.087	5.115	7.628	6.103
Buffered fold 2	2,112	4.897	4.979	7.740	5.782

Buffered fold 3	1,344	6.413	6.527	8.066	9.965
Buffered fold 4	4,416	6.483	6.428	9.058	9.341
Train Mafraq	4,608	4.211	4.285	4.536	5.961
Train JV	3,840	7.807	7.743	12.441	9.947

Table S6. Annual and Regional Paired Differences for Every Overall Comparison

Comparison	Year or region	Rows	ΔMAE	ΔRMSE
Temporal HGB	2024	4,224	+0.0959	+0.2074
Temporal HGB	2025	4,224	+0.2146	+0.0856
Temporal HGB	JV	3,840	+0.1749	+0.1538
Temporal HGB	Mafraq	4,608	+0.1356	+0.1425
Temporal ridge	2024	4,224	+0.0754	+0.0437
Temporal ridge	2025	4,224	+0.0577	+0.0098
Temporal ridge	JV	3,840	+0.0587	+0.0068
Temporal ridge	Mafraq	4,608	+0.0744	+0.0595
Buffered spatial	2024	4,224	+0.0400	+0.1389
Buffered spatial	2025	4,224	-0.0002	-0.1789
Buffered spatial	JV	3,840	-0.0889	-0.0887
Buffered spatial	Mafraq	4,608	+0.1287	+0.1077
Zero buffer	2024	4,224	+0.1147	+0.1316
Zero buffer	2025	4,224	-0.0570	-0.0955
Zero buffer	JV	3,840	+0.0094	+0.0250
Zero buffer	Mafraq	4,608	+0.0484	+0.0137
Mafraq → JV	2024	1,920	-1.3467	-1.7975
Mafraq → JV	2025	1,920	-0.1417	-0.3725
Mafraq → JV	JV	3,840	-0.7442	-1.1283
Mafraq → JV	Mafraq	0	—	—

JV → Mafraq	2024	2,304	-0.5534	-0.6972
JV → Mafraq	2025	2,304	-0.6335	-0.8827
JV → Mafraq	JV	0	—	—
JV → Mafraq	Mafraq	4,608	-0.5935	-0.7863

Table S7. Paired Differences Under the Two Prespecified Sensitivity Masks

Comparison	Mask	Rows	ΔMAE	ΔRMSE
Temporal HGB	Full NDVI support	7,810	+0.1490	+0.1352
Temporal HGB	Both map editions	6,336	+0.1710	+0.1627
Temporal ridge	Full NDVI support	7,810	+0.0656	+0.0237
Temporal ridge	Both map editions	6,336	+0.0603	+0.0270
Buffered spatial	Full NDVI support	7,810	+0.0231	-0.0328
Buffered spatial	Both map editions	6,336	+0.0186	-0.0125
Zero buffer	Full NDVI support	7,810	+0.0236	-0.0206
Zero buffer	Both map editions	6,336	+0.0439	+0.0388
Mafraq → JV	Full NDVI support	3,534	-0.7892	-1.1638
Mafraq → JV	Both map editions	2,784	-0.7195	-1.0820
JV → Mafraq	Full NDVI support	4,276	-0.6014	-0.8143
JV → Mafraq	Both map editions	3,552	-0.5771	-0.7307

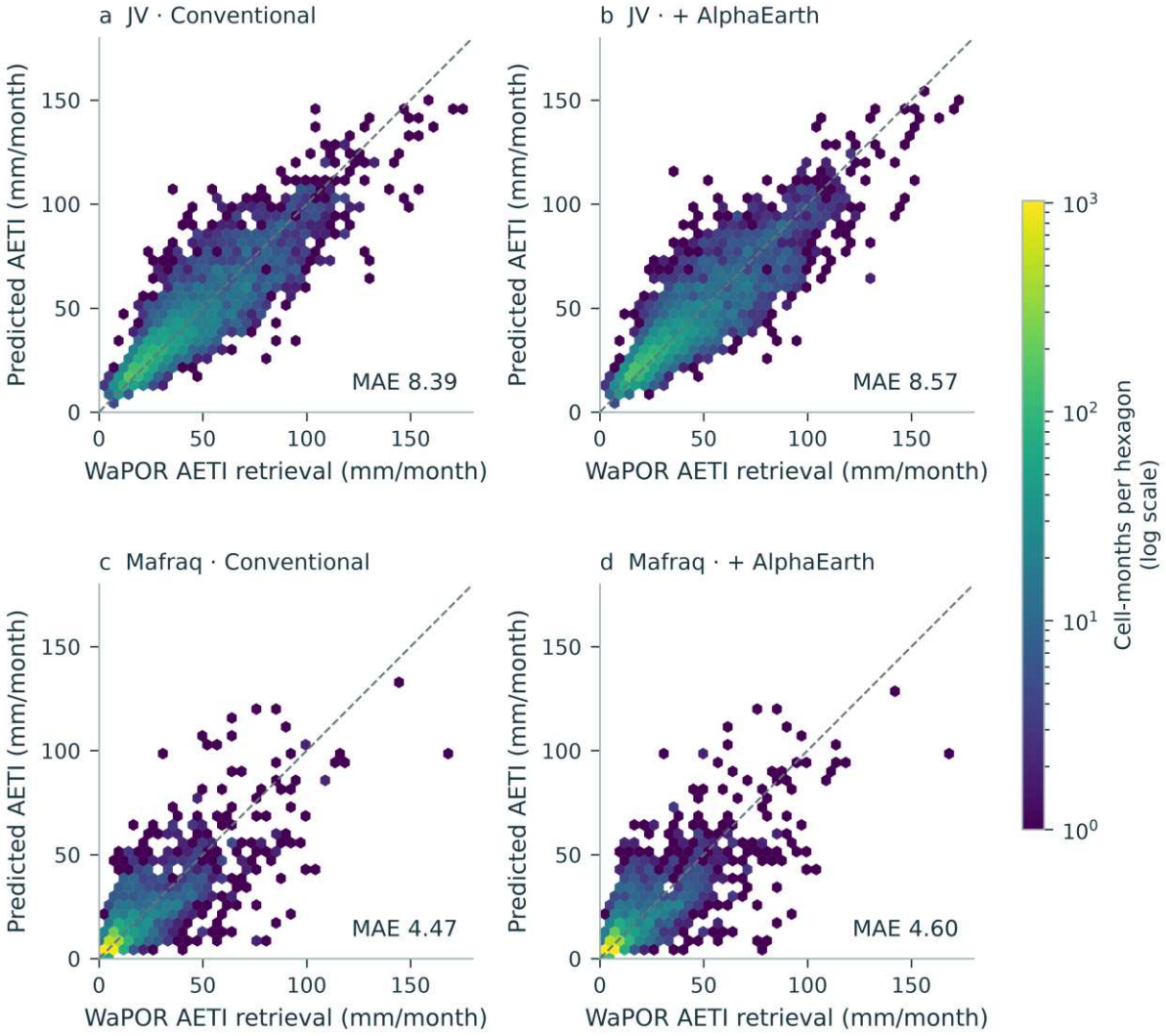


Figure S1. Prediction Density Relative to WaPOR AETI in Each Study Region

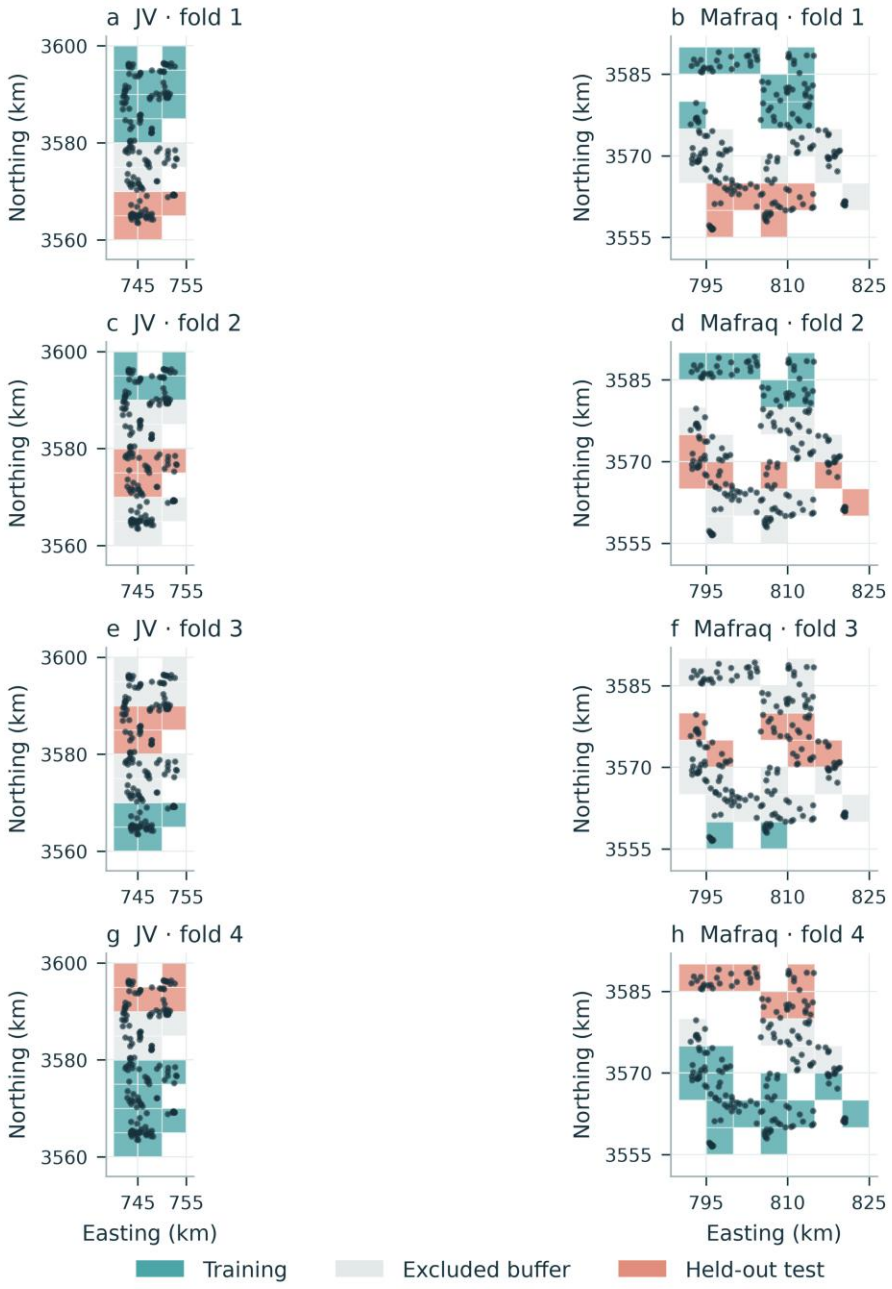


Figure S2. Training Blocks and Exclusions Across the Buffered Spatial Folds

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- GITHUBREPOVIDEO1.mp4