FROM ANNUAL TO IN-SEASON: UNCERTAINTY-AWARE DISTILLATION OF PRECOMPUTED GEOSPATIAL FOUNDATION MODEL EMBEDDINGS

Anonymous authors
Paper under double-blind review

ABSTRACT

Geospatial foundation models (GFMs) have demonstrated strong performance across a wide range of downstream tasks. Globally precomputed GFM embeddings are increasingly being released as ready-to-use geospatial layers. However, these embeddings are typically annual and retrospective, limiting their use for within-year tasks. Generating new embeddings can also be computationally expensive and impractical at large scales. This precludes their use for temporally-sensitive tasks like agricultural monitoring, where embeddings must be available and re-computed frequently within the growing season. We present an uncertainty-aware framework that distills annual embeddings into temporally evolving representations using partial satellite time series observations. Our model predicts a distribution over the annual embedding conditioned on observations available at a given date. As the year progresses, predicted embeddings converge to the annual representation while uncertainty decreases. We show that the resulting embeddings support in-season agricultural monitoring tasks including cropland classification and field boundary segmentation. Student embeddings achieve high cosine similarity to the annual teacher embedding even with one month of data, and downstream performance matches or exceeds the annual teacher in most tasks. Our method offers a lightweight model to extend annual GFM embedding products to in-season agricultural applications.

1 INTRODUCTION

Geospatial foundation models (GFMs) have emerged as a new paradigm for Earth observation, learning transferable representations from large-scale satellite imagery that generalize across diverse downstream tasks (e.g., Cong et al. (2022); Tseng et al. (2025); Clay (2025); Jakubik et al. (2025); Danish et al. (2026)). Rather than training task-specific models from scratch, practitioners can use pretrained GFMs to generate embeddings and train lightweight task-specific models on top of these representations (e.g., Zvonkov et al. (2025)). To make GFMs more accessible to users, many model developers are releasing globally precomputed embeddings as reusable geospatial products (Fang et al., 2026). Precomputed embedding products enable practitioners to integrate foundation model representations directly into downstream workflows without having to run costly models themselves (Brown et al., 2025; Clay, 2025; Herzog et al., 2026; Francis & Czerkawski, 2024; Feng et al., 2026a; Spheer, 2026).

However, precomputed GFM embeddings are fundamentally retrospective. Global embedding products are typically computed using a complete year of satellite observations and therefore cannot be produced until the full annual observation period is available (Francis & Czerkawski, 2024; Feng et al., 2026a; Clay, 2025; Brown et al., 2025). This creates a temporal mismatch for applications that require representations to be updated as new observations arrive (Nedungadi et al., 2026). For example, AlphaEarth Foundations (AEF) embeddings are currently available for 2017-2025 (Brown et al., 2025), which limits their use when observations from 2026 are required. In agricultural monitoring and food security analysis, for example, decisions need to be made during the growing season rather than after the year has ended (Nakalembe & Kerner, 2023; Anderson et al., 2026). Similar constraints arise in time-sensitive applications, such as disaster damage assessments.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

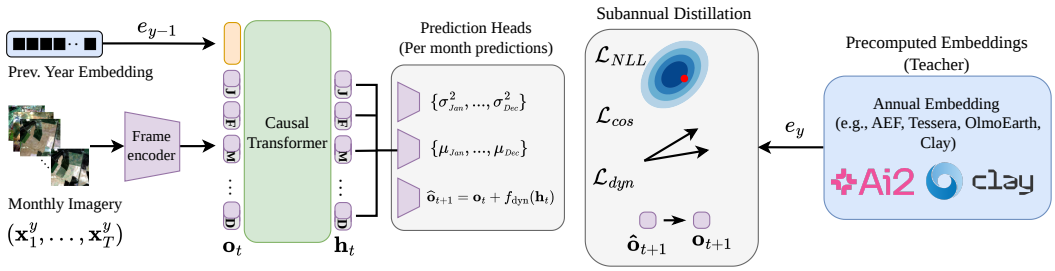


Figure 1: Overview of the proposed framework. A causal student predicts the annual GFM embedding and its uncertainty from partial in-season Sentinel-2 (S2) satellite observations.

A seemingly straightforward solution is to regenerate embeddings as new satellite observations arrive, but this is prohibitively expensive to do frequently and at large scale. Running GFMs over large geographic regions requires substantial computation (Fang et al., 2026; Muhawenayo et al., 2026), making repeated inference costly for downstream users (Klemmer et al.; Herzog et al., 2026). Recomputation is not possible for closed-source models like AEF (Brown et al., 2025). Some providers offer custom embeddings for user-defined temporal windows through an on-demand service or API (Tseng et al., 2026; Brown et al., 2025), but these services can be costly or unavailable for large scales or frequent computation. Thus, this creates a need for lightweight models that can update GFM embeddings as new satellite observations become available.

Motivated by this gap, we propose a new approach that estimates annual GFM representations during the year from partial satellite observations. We propose an uncertainty-aware temporal distillation framework (Fig. 1) in which a published annual GFM embedding serves as the teacher and a lightweight causal student learns to predict the teacher's annual embedding from partial Sentinel-2 satellite image time series. At each point in the season, the student predicts both an embedding and its uncertainty conditioned only on observations available up to that time. As additional observations arrive, the predicted representation converges toward the annual teacher embedding, while the uncertainty captures the residual ambiguity (see Fig. 2 for details of subannual to annual convergence).

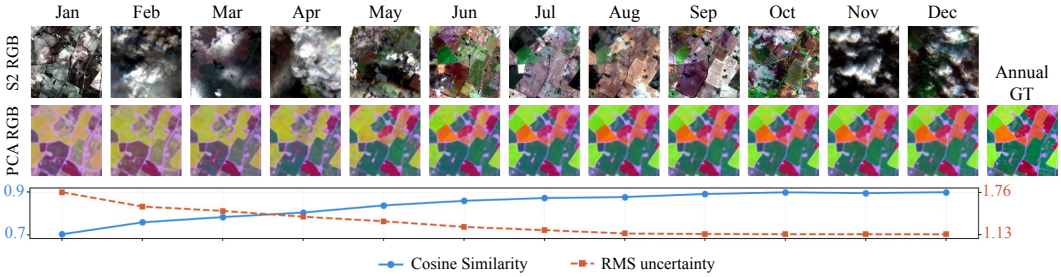


Figure 2: **Partial-year embeddings progressively converge to the annual representation.** Results are shown on a sample from the validation set in the distillation dataset.

We evaluated our approach on agricultural monitoring datasets: pixel-level cropland classification (Tárano et al., 2026; Wagner et al., 2025), cropland change detection, and agricultural field boundary segmentation using the Fields of The World (FTW) benchmark (Kerner et al., 2025). Results show that annual representations can be recovered early in the year and used for downstream agricultural tasks, while predicted uncertainty provides a useful measure of representation reliability.

Our study makes the following key contributions:

- We propose a new approach that formulates sub-annual GFM embedding prediction as a probabilistic temporal distillation problem with uncertainty estimation.
- Our lightweight student supports in-season mapping on standard hardware, updating embeddings and uncertainty for each new observation at about $4.8\,\text{km}^2$ per second on a 16-core CPU ($2.6\,\text{km}^2$ per second on 8 cores).

Together, these contributions enable researchers and practitioners to use annual embedding products for time-sensitive applications before the full year of observations is available. We open source the trained models, inference pipeline, and code to support reproducibility, deployment, and future work. Links are omitted during peer review.

2 Related Work

Computing sub-annual geospatial embeddings. Existing approaches provide several ways to obtain GFM embeddings before a complete year of observations is available (Lu et al., 2025). Open-weight models, such as Galileo, AnySat, OlmoEarth, and TESSERA, allow users to compute representations for user-defined temporal windows and regenerate them as new observations become available (Tseng et al., 2025; Astruc et al., 2025; Herzog et al., 2026; Feng et al., 2026a). This approach has been explored for in-season agricultural monitoring and other sub-annual Earth observation tasks (Shang et al., 2026; Athanasiadis, 2026; Nedungadi et al., 2026; Ma et al., 2026). Alternatively, hosted services can generate embeddings on demand for user-defined regions and time windows; e.g., OlmoEarth Studio supports custom exports over one to twelve monthly observation periods (Tseng et al., 2026; Herzog et al., 2026). This flexibility comes at the cost of accessibility, storage, and computing resources, requiring repeated embedding retrieval or model inference. Precomputed embedding products avoid this inference burden for downstream users, but are commonly released at fixed temporal resolutions (Fang et al., 2026; Brown et al., 2025; Feng et al., 2026a). Our work aims to preserve the accessibility of precomputed annual embedding products, using them as supervision to estimate annual representations from observations available sub-annually.

GFM distillation. Knowledge distillation is widely used to transfer knowledge from large teacher models to smaller students, primarily for model compression and computational efficiency (Hinton et al., 2015; Himeur et al., 2024). In Earth observation, distillation has been used both to transfer foundation-model knowledge to task-specific models (Gao et al., 2024) and to train smaller foundation models (Mendieta et al., 2023). Recent approaches have also directly distilled GFM representations. TESSERA v2 distills a large pixel-wise teacher into compact students and uses a Matryoshka objective to support multiple embedding dimensionalities (Feng et al., 2026b). This reduces model and representation size, but its distillation objective varies embedding dimensionality rather than the temporal availability of the representation. BetaEarth demonstrates that publicly released AEF embeddings can supervise an independent model without access to the underlying AEF model weights, and can aggregate scene-level predictions across multiple observations (Czerkawski, 2026; Brown et al., 2025). Our work builds on this broader direction of embedding distillation but focuses on temporal availability. We train a causal student to recover a fixed annual representation from progressively longer prefixes of in-season observations and estimate uncertainty at each cutoff to represent the ambiguity remaining under incomplete seasonal observations.

3 Method

We formulate in-season representation learning as temporal distillation (Hinton et al., 2015). A lightweight causal student predicts a distribution over the annual GFM embedding from available satellite observations, capturing uncertainty under incomplete seasonal information. An auxiliary dynamics objective encourages the student to learn changes between consecutive months.

3.1 Problem Setup

Let Φ be a frozen GFM that maps a full year of satellite observations at a given location to an annual embedding. For year y, we write $\mathbf{e}_y = \Phi(\mathbf{X}^y_{1:T}) \in \mathbb{R}^E$, where E is the embedding dimension and $T = 12$ is the number of months in the year. We write $\mathbf{X}^y = (\mathbf{x}^y_1, \ldots, \mathbf{x}^y_T)$, where each $\mathbf{x}^y_t \in \mathbb{R}^{C\times P\times P}$ is a monthly median composite of the Sentinel-2 scenes acquired in month t of year y, given as a C-band patch of $P \times P$ pixels centered on the target location. For simplicity, we omit the year index for current-year quantities, writing $\mathbf{e} = \mathbf{e}_y$, $\mathbf{X} = \mathbf{X}^y$, and $\mathbf{x}_t = \mathbf{x}^y_t$. During the year, only a prefix $\mathbf{X}_{\le t}$ of this series is available. At each step t, the student predicts a distribution over the annual embedding from the observations seen so far,

$$g_\theta : (\mathbf{X}_{\le t}, \mathbf{e}_{y-1}) \longmapsto q_\theta(\mathbf{e} \mid \mathbf{X}_{\le t}, \mathbf{e}_{y-1}), \tag{1}$$

and never accesses $\mathbf{x}_{>t}$. Here $\mathbf{e}_{y-1}$ is the teacher embedding at the same pixel from the previous year, a location-specific prior that is available before the season begins and therefore preserves the causal setting; we ablate the unconditioned variant in § A.3. Training uses precomputed embeddings for both signals: the current-year embedding $\mathbf{e}$ as supervision and $\mathbf{e}_{y-1}$ as conditioning.

Early in the season, available observations may not fully determine the annual representation. We therefore predict both an embedding and its uncertainty. Following probabilistic embedding approaches (Shi & Jain, 2019; Oh et al., 2019), we estimate uncertainty separately for each embedding dimension, allowing some components to be less certain than others. We model the dimensions independently rather than estimating correlations between them, which keeps the uncertainty prediction lightweight. Specifically, for each embedding dimension j,

$$e_j \sim \mathcal{N}\left(\mu_j^{(t)}, (\sigma_j^{(t)})^2\right), \qquad j = 1, \ldots, E. \tag{2}$$

Here, $\boldsymbol{\mu}^{(t)} \in \mathbb{R}^E$ is the predicted annual embedding, and $\boldsymbol{\sigma}^{(t)} \in \mathbb{R}^E$ contains its per-dimension uncertainty. As observations accumulate, we expect the prediction to approach the annual teacher embedding and uncertainty to decrease.

3.2 Student model

The student consists of a spatial encoder, a causal temporal encoder, embedding and uncertainty heads, and an auxiliary dynamics head (Fig. 1).

Per-frame spatial encoding. A shared convolutional encoder processes each monthly Sentinel-2 patch without spatial downsampling. Learned cross-attention pooling (Touvron et al., 2021) summarizes the feature map into one observation token per month, which we refer to as $\mathbf{o}_t$:

$$\mathbf{o}_t = \mathrm{pool}(f_{\mathrm{enc}}(\mathbf{x}_t)), \qquad t = 1, \ldots, T, \tag{3}$$

where f_{enc} is the spatial encoder. Pooling incorporates surrounding context into the target pixel's representation. We compare cross-attention pooling with center-pixel features and center-pixel features concatenated with mean-pooled context in § A.3.

Causal temporal encoding. The monthly tokens are processed by a Transformer with causal attention (Melnychuk et al., 2022), so the representation at month t attends only to months $1, \ldots, t$. The previous-year embedding $\mathbf{e}_{y-1}$ is projected into the student feature space by a learned linear layer followed by a GELU nonlinearity, then prepended to the sequence. The resulting temporal representation $\mathbf{h}_t$ summarizes both the previous-year prior and the observations through month t.

In-season embedding and uncertainty heads. A shared trunk maps $\mathbf{h}_t$ to two heads: one predicting the inseason embedding mean $\boldsymbol{\mu}^{(t)}$ the other the log-variance $\mathbf{s}^{(t)} = \log\left((\boldsymbol{\sigma}^{(t)})^2\right)$ (Kendall & Gal, 2017). Predicting log-variance improves numerical stability; we clamp its values to a fixed range and initialize the head to output $\mathbf{s}^{(t)} = \mathbf{0}$, corresponding to unit variance in each dimension.

Auxiliary seasonal dynamics head. The distillation model predicts the same annual embedding at each time step. This drives progressive recovery of the teacher representation, but does not require $\mathbf{h}_t$ to model how the scene changes between consecutive months. Inspired by next-latent prediction (Teoh et al., 2026), we add a next-month prediction task. A small dynamics head f_{dyn} takes $\mathbf{h}_t$ and predicts the change between the current and next month in the observation token:

$$\widehat{\mathbf{o}}_{t+1} = \mathbf{o}_t + f_{\mathrm{dyn}}(\mathbf{h}_t). \tag{4}$$

The current observation token $\mathbf{o}_t$ provides a persistence baseline, while $f_{\mathrm{dyn}}(\mathbf{h}_t)$ predicts the change needed to obtain the next-month token. Targets $\mathbf{o}_{t+1}^{\mathrm{tar}}$ come from a slowly updated copy of the spatial encoder and pooling module, following the momentum target-encoder in joint-embedding predictive architectures (Assran et al., 2023). Gradients do not propagate through this target encoder.

3.3 Training objective

Training requires no downstream labels. The published precomputed annual embedding $\mathbf{e}$ serves as the supervision target (Brown et al., 2025; Feng et al., 2026a). Because the temporal encoder

is causal, a single forward pass produces predictions at every cutoff, and all valid time steps $\mathcal{O}$ are supervised jointly. The training objective combines annual embedding recovery with the auxiliary seasonal dynamics task. For annual embedding recovery, we jointly learn the predicted embedding and its uncertainty using Gaussian negative log-likelihood $\mathcal{L}_{\text{NLL}}$ (Nix & Weigend, 1994; Kendall & Gal, 2017). We additionally use cosine distance $\mathcal{L}_{\cos}$ to encourage directional alignment with the annual teacher embedding. The seasonal dynamics loss $\mathcal{L}_{\text{dyn}}$ supervises next-observation prediction and encourages the temporal representation to capture changes between consecutive observations. These losses are defined as

$$\mathcal{L}_{\text{NLL}} = \frac{1}{|\mathcal{O}|} \sum_{t \in \mathcal{O}} \left\langle \tfrac{1}{2} \left[\exp(-\mathbf{s}^{(t)}) \odot (\mathbf{e} - \boldsymbol{\mu}^{(t)})^2 + \mathbf{s}^{(t)} \right] \right\rangle, \tag{5}$$

$$\mathcal{L}_{\cos} = \frac{1}{|\mathcal{O}|} \sum_{t \in \mathcal{O}} \left[1 - \cos(\boldsymbol{\mu}^{(t)}, \mathbf{e}) \right], \tag{6}$$

$$\mathcal{L}_{\text{dyn}} = \text{MSE}(\widehat{\mathbf{o}}_{t+1}, \mathbf{o}^{\text{tar}}_{t+1}) + \lambda_{\text{dyn,cos}} \left[1 - \cos(\widehat{\mathbf{o}}_{t+1}, \mathbf{o}^{\text{tar}}_{t+1}) \right]. \tag{7}$$

Here, $\langle \cdot \rangle$ denotes the average over embedding dimensions. The NLL penalizes embedding error according to predicted confidence while also penalizing unnecessarily large uncertainty. The complete objective combines the three losses as

$$\mathcal{L} = \mathcal{L}_{\text{NLL}} + \lambda_{\cos} \mathcal{L}_{\cos} + \lambda_{\text{dyn}} \mathcal{L}_{\text{dyn}}. \tag{8}$$

3.4 Within-year inference

At inference, the student processes observations causally as they become available. Given observations through month t and the previous-year embedding, a forward pass produces the in-season embedding $\boldsymbol{\mu}^{(t)}$ and its predicted standard deviation $\boldsymbol{\sigma}^{(t)}$. The predicted embedding can be used in place of the annual teacher embedding for downstream tasks. As new observations arrive, the sequence is extended without access to future observations. The auxiliary dynamics head is used only during training and is discarded at inference. Inference therefore requires only the spatial encoder, causal temporal encoder, and prediction heads for $\boldsymbol{\mu}^{(t)}$ and $\boldsymbol{\sigma}^{(t)}$. Implementation details and throughput measurements for the large-scale inference pipeline are provided in Appendix A.4.

4 Experimental Setup

4.1 Distillation dataset

We use AEF (Brown et al., 2025) as the primary teacher since this is the only multi-year globally-available embedding product. We separately distill TESSERA-v1 (Feng et al., 2026a) to compare transfer between embedding products. Unlike its global 2024 product, TESSERA's 2023 embeddings, used for previous-year conditioning, cover only 42.7% of our sampled locations. We evaluate missing-prior strategies in § A.3 and report TESSERA transfer results in § 5.5.

We sample 71,771 locations using locations from the CropHarvest dataset (Tseng et al., 2021) to guide sampling toward agriculturally relevant areas. CropHarvest labels are not used for supervision, only for sampling locations. Each example pairs a 2024 time series of up to 12 monthly median composites of Sentinel-2 L2A scenes with less than 40% scene cloud cover ($C = 12$ bands, 15×15 patches) with the co-located 2024 annual embedding as the distillation target ($\mathbf{e} \in \mathbb{R}^{64}$ for AEF, $\mathbf{e} \in \mathbb{R}^{128}$ for TESSERA) (Source et al., 2022; Brown et al., 2025; Feng et al., 2026a). Months without a valid Sentinel-2 composite are zero-filled and masked from attention, with loss computed only at valid cutoffs. We randomly split the samples into training (85%) and validation (15%) sets.

4.2 Evaluation

We evaluate the learned sub-annual representations in two complementary ways. First, we measure embedding fidelity by comparing student embeddings at monthly cutoffs with the annual teacher.

Second, we measure downstream performance on classification, change detection and segmentation tasks. Together, these evaluations measure how quickly the student recovers the annual representation and how early it becomes useful for downstream applications.

Classification and change-detection datasets. Cropland classification is a binary task that distinguishes crop from non-crop points. We evaluate it in Sudan and Ukraine (Tárano et al., 2026; Wagner et al., 2025), where fixed reference points are labeled over four seasons. The datasets contain 1,636 points in Sudan (2022–2025) and 2,000 in Ukraine (2021–2024). For each point, we extract the Sentinel-2 time series and generate student embeddings at each monthly cutoff.

Cropland maps are also used to track change, such as fields abandoned during conflict or left uncultivated in a drought (Kerner et al., 2024; Nakalembe et al., 2021). We thus separately evaluate change using points labeled in consecutive seasons, giving 4,556 pairs in Sudan and 6,000 in Ukraine, of which 17% and 13% changed class, respectively.

Segmentation dataset. We evaluate field boundary segmentation using Fields of The World (FTW) (Kerner et al., 2025), which contains over 1.5 million validated field polygons across 24 countries. Since FTW provides bi-temporal RGB-NIR Sentinel-2 imagery, we exported full-year, 12-band monthly Sentinel-2 observations for each chip to generate student embeddings at each cutoff and the corresponding annual teacher embeddings. We use predefined per-country splits and exclude presence-only countries (Brazil, India, Kenya, and Rwanda) that lack background labels.

Downstream models. For cropland classification, we use an ℓ_2-regularized logistic regression probe with standardized features and class-balanced weights. At each cutoff t, we evaluate the current prediction $\boldsymbol{\mu}_t$, its uncertainty-augmented representation $[\boldsymbol{\mu}_t, \boldsymbol{\sigma}_t]$, and all predictions so far $\boldsymbol{\mu}_{1..t}$. Baselines include the previous-year and annual AEF embeddings, center-pixel Sentinel-2 reflectance, and monthly Dynamic World (Brown et al., 2022). We use $0.2°$ spatial-block cross-validation (5 folds, 3 repeats) to separate fields, seasons, and neighboring points between training and test. We report macro F1 with 95% spatial-block bootstrap intervals and align monthly cutoffs with the local crop calendar. Change detection is evaluated by comparing cropland maps from consecutive seasons.

For field boundary segmentation, we train a 1×1 convolutional probe on the teacher and student embeddings to predict background, field, and boundary classes. Following PRUE (Muhawenayo et al., 2026), we train a single model jointly across all countries using the same optimizer, learning-rate schedule, and class weights. Models are trained for 50 epochs, retaining the checkpoint with the lowest validation loss. We select a learning rate of 3×10^{-2} on the teacher validation set and use it for all student embeddings. Following the FTW and PRUE evaluation protocols (Kerner et al., 2025; Muhawenayo et al., 2026), we report per-country pixel IoU for the field class.

5 Results and Analysis

Our results show three main findings. First, the student recovers the annual teacher representation early, with larger gains when the previous-year embedding is a poor prior (Fig. 3). Second, the recovered embeddings capture current-season information, improving agricultural mapping as the season progresses, particularly for fields whose cultivation status changed (Figs. 4 and 5). Third, predicted uncertainty decreases through the season and identifies less reliable embeddings, although less effectively than downstream classifier confidence (Fig. 3; §5.2). These findings also hold for field boundary delineation (Fig. 6) and when distilling TESSERA embeddings instead of AEF (§5.5).

5.1 Annual representations are recovered progressively

Much of the annual representation is recovered early. In both regions, the student is already highly similar to the annual AEF embedding at the first monthly cutoff (blue, left axis of Fig. 3). January cosine similarity is 0.939 in Sudan and 0.912 in Ukraine, showing that much of the annual representation can be recovered from limited current-year observations together with the previous-year prior. Recovery continues as additional observations become available, reaching 0.955 in Sudan and 0.950 in Ukraine by December.

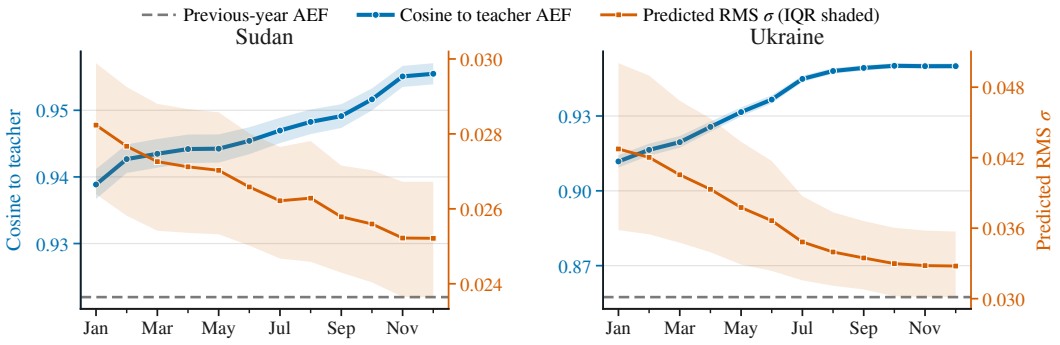


Figure 3: **Embedding recovery and predicted uncertainty through the year in Sudan and Ukraine.** Blue, left axis, is cosine similarity between the student embedding and the annual teacher AEF embedding at each causal monthly cutoff, with a 95% bootstrap interval. Orange, right axis, is the predicted RMS uncertainty with its p25 to p75 spread across points. The grey dashed line is the cosine of the previous-year AEF to the same teacher.

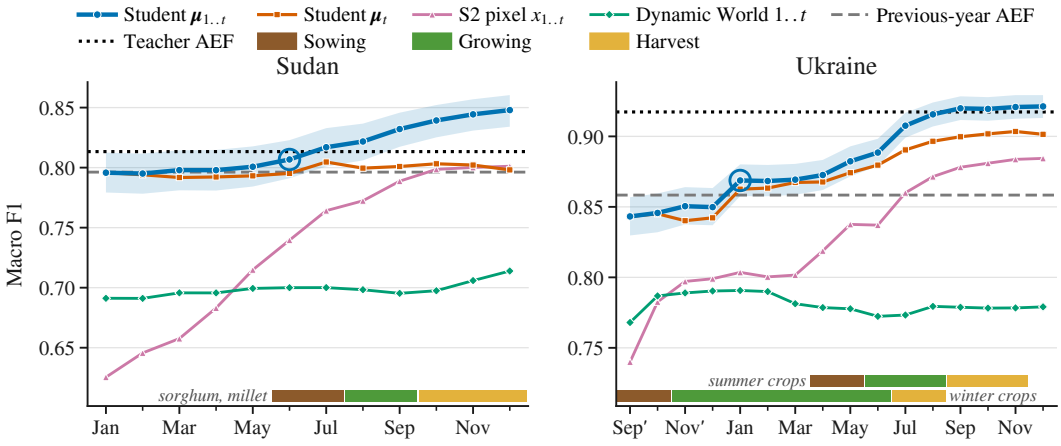


Figure 4: **Cropland classification through the growing season in Sudan and Ukraine.** Macro F1 of a logistic regression probe at each causal monthly cutoff t (the last month observed; $'$ marks the previous calendar year), using the student's latest prediction $\boldsymbol{\mu}_t$ or all predictions so far $\boldsymbol{\mu}_{1..t}$. The band shows the 95% spatial-block bootstrap interval for $\boldsymbol{\mu}_{1..t}$, and the ring marks the first cutoff whose interval is above the previous-year AEF. Dynamic World uses all months observed so far.

Current-year observations matter most where the previous year is a poor prior. In Ukraine, the previous year is a weak prior at 0.857, and the student improves on it throughout the season, from +0.054 in January to +0.093 in December. In Sudan, the previous year is stronger at 0.922, and the gain is correspondingly smaller, from +0.017 to +0.034 (Fig. 3). Both differences are paired on the same resampled spatial blocks and are outside their 95% intervals at every cutoff, suggesting that current-year observations provide the greatest benefit when conditions differ from the previous year.

Uncertainty decreases as observations accumulate. Predicted uncertainty decreases from January to December, from 0.043 to 0.033 in Ukraine and from 0.028 to 0.025 in Sudan (Fig. 3). It also tracks embedding error: the Spearman correlation between predicted σ and L_2 reconstruction error is 0.29 in Sudan and 0.57 in Ukraine, while the most uncertain quartile has 1.30× and 1.42× the error of the least uncertain quartile. However, December 1σ coverage is only 54.2% and 61.8%, respectively, below the nominal 68%. We therefore interpret uncertainty as a relative confidence measure rather than a calibrated probability interval.

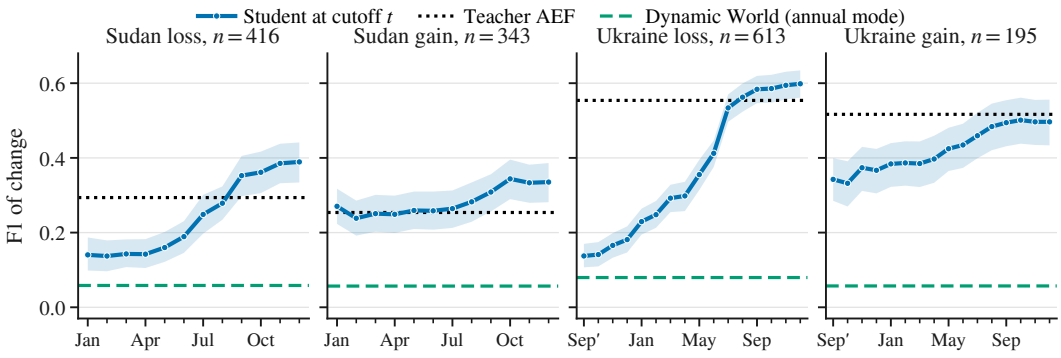


Figure 5: **F1 score on fields whose cultivation status changed since the previous season.** F1 score of the student $\boldsymbol{\mu}_{1..t}$ probe at each cutoff, with 95% spatial-block bootstrap intervals. For fields that left cultivation, the previous-year AEF reflects last season's status and is therefore mostly incorrect. The blue dashed line restricts Sudan to points labelled with confidence in both seasons.

5.2 Recovered embeddings capture current-season crop changes

Accuracy improves as the growing season progresses. Fig. 4 shows that in Sudan, before planting, the student performs similarly to the previous-year AEF. After planting, it improves over the previous-year AEF as crops develop, from June in Sudan and January in Ukraine after winter crop green-up (Fig. 4). It matches the AEF oracle from August in Ukraine and exceeds it from September in Sudan ($+0.019$ [$+0.008$, $+0.029$]), months before the annual embedding is available. By December, it is 0.052 and 0.063 above the previous-year AEF in Sudan and Ukraine, respectively.

Using every monthly prediction matters. A probe given only the latest prediction $\boldsymbol{\mu}_t$ never matches the annual AEF oracle. Giving the same linear probe all predictions accumulated so far, $\boldsymbol{\mu}_{1..t}$, adds 0.050 [0.039, 0.060] in Sudan and 0.020 [0.014, 0.026] in Ukraine by December. This is consistent with prior work showing the importance of temporal information for agricultural mapping (Sainte Fare Garnot & Landrieu, 2021; Butsko et al., 2025). Adding predicted $\boldsymbol{\sigma}_t$ to $\boldsymbol{\mu}_t$ also helps in Sudan ($+0.015$ [$+0.009$, $+0.022$]), but not significantly in Ukraine.

Existing cropland maps score well below the student. Dynamic World (Brown et al., 2022), using all images observed so far, reaches at most 0.71 in Sudan and 0.79 in Ukraine, and stays nearly flat as the season progresses, ending about 0.13 below the student in both regions (Fig. 4). Sentinel-2 reflectance at the centre pixel improves over the same months but remains below the student throughout.

Predicted uncertainty is a weak warning signal. Excluding the 20% of points with the highest predicted uncertainty $\boldsymbol{\sigma}$ raises end-of-season Macro F1 by 0.7–0.8 points, compared with 5–6 points when excluding those with lowest probe confidence. Unlike probe confidence, $\boldsymbol{\sigma}$ is available without downstream labels and can provide an early warning of unreliable embeddings.

5.3 Detecting cropland change within the season

The student detects fields whose cultivation status changed. A previous-year map cannot reveal change, and an annual map reveals it only after the season ends. On fields that stopped being cultivated, the previous-year AEF is correct for only 29% in Sudan and 15% in Ukraine, whereas the student's F1 score rises through the season to 39% and 60% by December, above the annual AEF oracle (29% and 55%; Fig. 5). The gap holds on Sudanese points labelled with confidence in both seasons (39% against 27%). On newly cultivated fields, the student improves in Ukraine (34% to 50%) but not in Sudan, and on unchanged fields all embeddings reach 85–97%. Comparing consecutive seasonal maps without new labels, the student's December maps detect change with an F1 of 0.42 in Sudan and 0.58 in Ukraine, outperforming consecutive AEF oracle maps in both regions.

Table 1: **Cropland macro F1 through the season with TESSERA.** Same protocol as for AEF. The student is conditioned on the previous-year TESSERA embedding and evaluated where it is available; annual references are evaluated only on the point-seasons where they are available (n). Brackets are 95% intervals. Best in-season value per column in bold.

	Sudan			Ukraine		
	Mar (pre)	Aug (mid)	Dec (end)	Dec′ (winter)	May (spring)	Dec (end)
Raw S2 pixel $x_{1..t}$	0.633	0.763	0.786	0.799	0.838	0.885
Dynamic World, t	0.687	0.563	0.680	0.545	0.730	0.496
Dynamic World, $1..t$	0.679	0.710	0.718	0.790	0.777	0.779
Student $\boldsymbol{\mu}_t$	**0.769**	0.805	0.831	0.858	0.877	0.911
Student $[\boldsymbol{\mu}_t, \boldsymbol{\sigma}_t]$	0.768	**0.809**	**0.835**	0.860	0.879	0.913
Student $\boldsymbol{\mu}_{1..t}$	**0.769**	0.790	0.833 [0.814, 0.851]	**0.862**	**0.881**	**0.923** [0.915, 0.930]
Prev.-year TESSERA	0.790 [0.765, 0.815], $n = 1{,}796$			0.864 [0.852, 0.876], $n = 7{,}992$		
Annual TESSERA	0.770 [0.727, 0.810], $n = 504$			0.927 [0.919, 0.934], $n = 7{,}992$		

5.4 Field boundary delineation

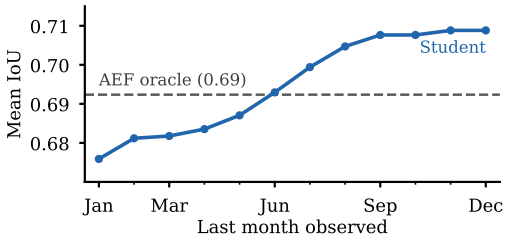


Figure 6: **Mean FTW field IoU** of monthly student embeddings against the annual AEF oracle, over the 17 countries with at least 30 test chips.

Field delineation improves as the season progresses. Mean field IoU of the student increases steadily through the season, from 0.68 in January to 0.71 by September, and remains stable thereafter (Fig. 6). It reaches the annual AEF oracle (0.69) in June and exceeds it from July onward, months before the annual embedding becomes available. Results vary across countries, with the largest gain in the Netherlands and the student remaining below the oracle in Luxembourg. We present per-country results in Appendix B.1 (Table 4).

Uncertainty tracks embedding error on FTW. Across FTW countries, the student reaches 0.93–0.97 cosine similarity to the annual teacher, and uncertainty correlates with embedding error in most countries (Pearson r up to 0.85; Appendix Table 3). However, 1σ coverage (39–66%) is below the nominal 68.3%, supporting its use as a relative rather than calibrated measure of reliability.

5.5 The distillation framework transfers across embedding products

The same temporal behavior appears with TESSERA. Because the student requires only published annual embeddings for supervision, the framework applies to other embedding products without modification. Distilling TESSERA instead of AEF shows the same trends discussed in §5.1, 5.3 and 5.4. Similarity to the annual teacher increases through the season, from 0.863 to 0.899 in Sudan and from 0.830 to 0.883 in Ukraine, and predicted uncertainty decreases and tracks embedding error (Appendix Fig. 11). Downstream results also follow the AEF pattern: cropland F1 improves as the season progresses, reaching 0.833 in Sudan and 0.923 in Ukraine by December, above the previous-year TESSERA (0.790 and 0.864) and comparable to the annual TESSERA teacher (Table 1). Change detection and field boundary delineation show similar trends (Appendix B.3).

6 Conclusion

We presented a lightweight temporal distillation framework for extending annual precomputed GFM embeddings to in-season applications. Across two embedding products and agricultural tasks, the student produces useful representations before annual products become available, while providing uncertainty estimates of their reliability. Results show that published annual embeddings can serve not only as static products, but also as supervision for models that update representations as observations arrive. This offers a practical way to use GFM embedding products in time-sensitive applications without access to or repeated inference with the underlying foundation model.

AI USE STATEMENT

In this work, we used generative AI tools to assist with interpreting experimental results and implementing parts of the method and data pipeline. We did not use generative AI tools to generate data, formulate mathematical claims or hypotheses, design experiments, or clean datasets. Translation and qualitative analysis were not applicable to this work. Additionally, we used generative AI tools to write and edit code, format tables, and draft and edit parts of the manuscript for clarity and length. We reviewed all AI-assisted work. AI-generated code was run and tested by the authors, and AI-assisted interpretations were verified against the underlying experimental results. We take responsibility for the final content of this work, including text, claims, code, and artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

We have taken several steps to make this work reproducible. The model architecture and training objective are specified in §3, including input and output shapes. The distillation dataset construction is described in §4.1, including sampling locations, monthly compositing, teacher embeddings, and train–validation splits. Both teacher embeddings and sampling locations are from public sources. Downstream evaluation protocols are described in §4.2 and all downstream datasets are from published benchmarks. Ablations in §A.3 modify one setting of a shared reference configuration and report seed-to-seed variation. Source code for data export, model training, evaluation, and all ablation configurations will be released upon publication. Links are omitted during peer review.

REFERENCES

Weston Anderson, Inbal Becker-Reshef, Nicholas Bodanac, Bia Carneiro, Dave English, Andrew Hoell, Tim Hoffine, YanYan Liu, Mélissande Machefer, Michele Meroni, et al. Responsible use of artificial intelligence and machine learning for food security early warning systems. *Nature Food*, pp. 1–4, 2026.

Jason Ansel, Edward Yang, Horace He, Natalia Gimelshein, Animesh Jain, Michael Voznesensky, Bin Bao, Peter Bell, David Berard, Evgeni Burovski, Geeta Chauhan, Anjali Chourdia, Will Constable, Alban Desmaison, Zachary DeVito, Elias Ellison, Will Feng, Jiong Gong, Michael Gschwind, Brian Hirsh, Sherlock Huang, Kshiteej Kalambarkar, Laurent Kirsch, Michael Lazos, Mario Lezcano, Yanbo Liang, Jason Liang, Yinghai Lu, C. K. Luk, Bert Maher, Yunjie Pan, Christian Puhrsch, Matthias Reso, Mark Saroufim, Marcos Yukio Siraichi, Helen Suk, Shunting Zhang, Michael Suo, Phil Tillet, Xu Zhao, Eikan Wang, Keren Zhou, Richard Zou, Xiaodong Wang, Ajit Mathews, William Wen, Gregory Chanan, Peng Wu, and Soumith Chintala. Pytorch 2: Faster machine learning through dynamic python bytecode transformation and graph compilation. In *Proceedings of the 29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 2*, ASPLOS '24, pp. 929–947, New York, NY, USA, 2024. Association for Computing Machinery. ISBN 9798400703850. doi: 10.1145/3620665.3640366. URL https://doi.org/10.1145/3620665.3640366.

Mahmoud Assran, Quentin Duval, Ishan Misra, Piotr Bojanowski, Pascal Vincent, Michael Rabbat, Yann LeCun, and Nicolas Ballas. Self-supervised learning from images with a joint-embedding predictive architecture. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 15619–15629, June 2023.

Guillaume Astruc, Nicolas Gonthier, Clément Mallet, and Loic Landrieu. AnySat: One Earth Observation Model for Many Resolutions, Scales, and Modalities. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 19530–19540, June 2025.

Ioannis N. Athanasiadis. Foundation models for agricultural sciences - challenges and opportunities. In *ICLR 2026 Workshop on Foundation Models for Science: Real-World Impact and Science-First Design*, 2026. URL https://openreview.net/forum?id=2gewa9tg5p.

Christopher F. Brown, Steven P. Brumby, Brookie Guzder-Williams, Tanya Birch, Samantha Brooks Hyde, Joseph Mazzariello, Wanda Czerwinski, Valerie J. Pasquarella, Robert Haertel, Simon

Ilyushchenko, Kurt Schwehr, Mikaela Weisse, Fred Stolle, Craig Hanson, Oliver Guinan, Rebecca Moore, and Alexander M. Tait. Dynamic World, Near real-time global 10 m land use land cover mapping. *Scientific Data*, 9(1):251, June 2022. ISSN 2052-4463. doi: 10.1038/s41597-022-01307-4.

Christopher F Brown, Michal R Kazmierski, Valerie J Pasquarella, William J Rucklidge, Masha Samsikova, Chenhui Zhang, Evan Shelhamer, Estefania Lahera, Olivia Wiles, Simon Ilyushchenko, et al. Alphaearth foundations: An embedding field model for accurate and efficient global mapping from sparse label data. *arXiv preprint arXiv:2507.22291*, 2025.

Christina Butsko, Gabriel Tseng, Kristof Van Tricht, Giorgia Milli, David Rolnick, Ruben Cartuyvels, Inbal Becker-Reshef, Zoltan Szantoi, and Hannah Kerner. Deploying geospatial foundation models in the real world: Lessons from worldcereal. In Nicolas Audebert, Hossein Azizpour, Valentin Barrière, Javiera Castillo Navarro, Mikolaj Czerkawski, Heng Fang, Alistair Francis, Valerio Marsocci, Andrea Nascetti, and Ritu Yadav (eds.), *Proceedings of The TerraBytes ICML Workshop: Towards global datasets and models for Earth Observation*, volume 292 of *Proceedings of Machine Learning Research*, pp. 13–31. PMLR, 19 Jul 2025. URL https://proceedings.mlr.press/v292/butsko25a.html.

Clay. The Clay Foundation Model - An open source AI model and interface for Earth, 2025. URL https://github.com/Clay-foundation/model. Accessed: 2025-11-13.

Yezhen Cong, Samar Khanna, Chenlin Meng, Patrick Liu, Erik Rozi, Yutong He, Marshall Burke, David B. Lobell, and Stefano Ermon. SatMAE: Pre-training transformers for temporal and multi-spectral satellite imagery. In Alice H. Oh, Alekh Agarwal, Danielle Belgrave, and Kyunghyun Cho (eds.), *Advances in Neural Information Processing Systems*, 2022. URL https://openreview.net/forum?id=WBhqzpF6KYH.

Mikolaj Czerkawski. Betaearth: Emulating closed-source earth observation models through their public embeddings. https://github.com/asterisk-labs/beta-earth, 2026. GitHub repository.

Muhammad Sohail Danish, Muhammad Akhtar Munir, Syed Roshaan Ali Shah, Muhammad Haris Khan, Rao Muhammad Anwer, Jorma Laaksonen, Fahad Shahbaz Khan, and Salman Khan. TerraFM: A scalable foundation model for unified multisensor earth observation. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=cBxuzdUDNx.

Heng Fang, Adam J. Stewart, Isaac Corley, Xiao Xiang Zhu, and Hossein Azizpour. Earth embeddings as products: Taxonomy, ecosystem, and standardized access, 2026. URL https://arxiv.org/abs/2601.13134.

Zhengpeng Feng, Clement Atzberger, Sadiq Jaffer, Jovana Knezevic, Silja Sormunen, Robin Young, Madeline C Lisaius, Markus Immitzer, Toby Jackson, James Ball, et al. Tessera: Temporal embeddings of surface spectra for earth representation and analysis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 34818–34831, 2026a.

Zhengpeng Feng, Sadiq Jaffer, Ira Shokar, Jovana Knezevic, Mark Elvers, Clement Atzberger, Robin Young, Aneesh Naik, Niall Robinson, Andrew Blake, et al. Tessera v2: Scaling pixel-wise earth foundation models. *arXiv preprint arXiv:2607.03949*, 2026b.

Alistair Francis and Mikolaj Czerkawski. Major tom: Expandable datasets for earth observation. In *IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium*, pp. 2935–2940, 2024. doi: 10.1109/IGARSS53475.2024.10640760.

Tianyi Gao, Wei Ao, Xing-ao Wang, Yuanhao Zhao, Ping Ma, Mengjie Xie, Hang Fu, Jinchang Ren, and Zhi Gao. Enrich, distill and fuse: Generalized few-shot semantic segmentation in remote sensing leveraging foundation model's assistance. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 2771–2780, 2024. doi: 10.1109/CVPRW63382.2024.00283.

Alessandro Giusti, Dan C. Cireşan, Jonathan Masci, Luca M. Gambardella, and Jürgen Schmidhuber. Fast image scanning with deep max-pooling convolutional neural networks. In *2013 IEEE International Conference on Image Processing*, pp. 4034–4038, 2013. doi: 10.1109/ICIP.2013.6738831.

Henry Herzog, Favyen Bastani, Yawen Zhang, Gabriel Tseng, Joseph Redmon, Hadrien Sablon, Ryan Park, Jacob Morrison, Alexandra Buraczynski, Karen Farley, et al. Olmoearth: Stable latent image modeling for multimodal earth observation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 34806–34817, 2026.

Yassine Himeur, Nour Aburaed, Omar Elharrouss, Iraklis Varlamis, Shadi Atalla, Wathiq Mansoor, and Hussain Al Ahmad. Applications of knowledge distillation in remote sensing: A survey, 2024. URL https://arxiv.org/abs/2409.12111.

Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network, 2015. URL https://arxiv.org/abs/1503.02531.

Johannes Jakubik, Felix Yang, Benedikt Blumenstiel, Erik Scheurer, Rocco Sedona, Stefano Maurogiovanni, Jente Bosmans, Nikolaos Dionelis, Valerio Marsocci, Niklas Kopp, et al. Terramind: Large-scale generative multimodality for earth observation. *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2025.

Alex Kendall and Yarin Gal. What uncertainties do we need in bayesian deep learning for computer vision? In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, pp. 5580–5590, Red Hook, NY, USA, 2017. Curran Associates Inc. ISBN 9781510860964.

Hannah Kerner, Snehal Chaudhari, Aninda Ghosh, Caleb Robinson, Adeel Ahmad, Eddie Choi, Nathan Jacobs, Chris Holmes, Matthias Mohr, Rahul Dodhia, Juan M Lavista Ferres, and Jennifer Marcus. Fields of the world: A machine learning benchmark dataset for global agricultural field boundary segmentation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(27): 28151–28159, Apr. 2025. doi: 10.1609/aaai.v39i27.35034. URL https://ojs.aaai.org/index.php/AAAI/article/view/35034.

Hannah R. Kerner, Catherine Nakalembe, Benjamin Yeh, Ivan Zvonkov, Sergii Skakun, Inbal Becker-Reshef, and Amy McNally. Satellite data shows resilience of tigrayan farmers in crop cultivation during civil war. *Science of Remote Sensing*, 10:100140, December 2024. ISSN 2666-0172. doi: 10.1016/j.srs.2024.100140. URL http://dx.doi.org/10.1016/j.srs.2024.100140.

Konstantin Klemmer, Esther Rolf, Marc Rußwurm, Gustau Camps-Valls, Mikolaj Czerkawski, Stefano Ermon, Alistair Francis, Nathan Jacobs, Hannah Kerner, Lester Mackey, Gengchen Mai, Oisin Mac Aodha, Markus Reichstein, Caleb Robinson, David Rolnick, Evan Shelhamer, Vincent Sitzmann, Devis Tuia, and Xiaoxiang Zhu. Earth embeddings: Towards ai-centric representations of our planet. URL https://api.semanticscholar.org/CorpusID:283741197.

Siqi Lu, Junlin Guo, James R. Zimmer-Dauphinee, Jordan M. Nieusma, Xiao Wang, Parker VanValkenburgh, Steven A. Wernke, and Yuankai Huo. Vision foundation models in remote sensing: A survey. *IEEE Geoscience and Remote Sensing Magazine*, 13(3):190–215, 2025. doi: 10.1109/MGRS.2025.3541952.

Yuchi Ma, Yawen Shen, Anu Swatantran, and David B. Lobell. Harvesting alphaearth: Benchmarking the geospatial foundation model for agricultural downstream tasks. *International Journal of Applied Earth Observation and Geoinformation*, 149:105258, 2026. ISSN 1569-8432. doi: https://doi.org/10.1016/j.jag.2026.105258. URL https://www.sciencedirect.com/science/article/pii/S1569843226001743.

Valentyn Melnychuk, Dennis Frauen, and Stefan Feuerriegel. Causal transformer for estimating counterfactual outcomes. In *International Conference on Machine Learning*, pp. 15545–15562. PMLR, 2022.

Matías Mendieta, Boran Han, Xingjian Shi, Yi Zhu, and Chen Chen. Towards geospatial foundation models via continual pretraining. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 16806–16816, 2023.

Paulius Micikevicius, Sharan Narang, Jonah Alben, Gregory Diamos, Erich Elsen, David Garcia, Boris Ginsburg, Michael Houston, Oleksii Kuchaiev, Ganesh Venkatesh, and Hao Wu. Mixed precision training. In *International Conference on Learning Representations*, 2018. URL https://openreview.net/forum?id=r1gs9JgRZ.

Maxim Milakov and Natalia Gimelshein. Online normalizer calculation for softmax, 2018. URL https://arxiv.org/abs/1805.02867.

Gedeon Muhawenayo, Caleb Robinson, Subash Khanal, Zhanpei Fang, Isaac Corley, Alexander Wollam, Tianyi Gao, Leonard Strnad, Ryan Avery, Lyndon Estes, et al. Prue: A practical recipe for field boundary segmentation at scale. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2026.

Catherine Nakalembe and Hannah Kerner. Considerations for AI-EO for Agriculture in Sub-Saharan Africa. *Environmental Research Letters*, 18(4):041002, mar 2023. doi: 10.1088/1748-9326/acc476. URL https://doi.org/10.1088/1748-9326/acc476.

Catherine Nakalembe, Inbal Becker-Reshef, Rogerio Bonifacio, Guangxiao Hu, Michael Laurence Humber, Christina Jade Justice, John Keniston, Kenneth Mwangi, Felix Rembold, Shraddhanand Shukla, Ferdinando Urbano, Alyssa Kathleen Whitcraft, Yanyun Li, Mario Zappacosta, Ian Jarvis, and Antonio Sanchez. A review of satellite-based global agricultural monitoring systems available for africa. *Global Food Security*, 29:100543, 2021. ISSN 2211-9124. doi: https://doi.org/10.1016/j.gfs.2021.100543. URL https://www.sciencedirect.com/science/article/pii/S2211912421000523.

Vishal Nedungadi, Xingguo Xiong, Marc Rußwurm, and Ioannis N. Athanasiadis. Foundation models meet agriculture: Challenges beyond pretraining, 2026. URL https://arxiv.org/abs/2608.30392.

D.A. Nix and A.S. Weigend. Estimating the mean and variance of the target probability distribution. In *Proceedings of 1994 IEEE International Conference on Neural Networks (ICNN'94)*, volume 1, pp. 55–60 vol.1, 1994. doi: 10.1109/ICNN.1994.374138.

Seong Joon Oh, Andrew C. Gallagher, Kevin P. Murphy, Florian Schroff, Jiyan Pan, and Joseph Roth. Modeling uncertainty with hedged instance embeddings. In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=r1xQQhAqKX.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241. Springer, 2015.

Vivien Sainte Fare Garnot and Loic Landrieu. Panoptic segmentation of satellite image time series with convolutional temporal attention networks. *ICCV*, 2021.

Osman Semih Kayhan and Jan C. van Gemert. On translation invariance in cnns: Convolutional layers can exploit absolute spatial location. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 14262–14273, 2020. doi: 10.1109/CVPR42600.2020.01428.

Pierre Sermanet, David Eigen, Xiang Zhang, Michael Mathieu, Rob Fergus, and Yann LeCun. Overfeat: Integrated recognition, localization and detection using convolutional networks. In *International Conference on Learning Representations (ICLR)*, 2014.

Zhuocheng Shang, Sanmay Das, and Ahmed Eldawy. Benchmarking geospatial foundation models for agriculture applications, 2026. URL https://arxiv.org/abs/2606.29664.

Evan Shelhamer, Jonathan Long, and Trevor Darrell. Fully convolutional networks for semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(4):640–651, 2017. doi: 10.1109/TPAMI.2016.2572683.

Yichun Shi and Anil Jain. Probabilistic face embeddings. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 6901–6910, 2019. doi: 10.1109/ICCV.2019.00700.

Microsoft Open Source, Matt McFarland, Rob Emanuele, Dan Morris, and Tom Augspurger. microsoft/planetarycomputer: October 2022, October 2022. URL https://doi.org/10.5281/zenodo.7261897.

Spheer. Spheer fm embeddings, 2026. URL https://huggingface.co/datasets/spheer/spheer-fm-embeddings.

Jayden Teoh, Manan Tomar, Kwangjun Ahn, Edward S. Hu, Tim Pearce, Pratyusha Sharma, Akshay Krishnamurthy, Riashat Islam, Alex Lamb, and John Langford. Next-latent prediction transformers learn compact world models, 2026. URL https://arxiv.org/abs/2511.05963.

Philippe Tillet, H. T. Kung, and David Cox. Triton: an intermediate language and compiler for tiled neural network computations. In *Proceedings of the 3rd ACM SIGPLAN International Workshop on Machine Learning and Programming Languages*, MAPL 2019, pp. 10–19, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450367196. doi: 10.1145/3315508.3329973. URL https://doi.org/10.1145/3315508.3329973.

Hugo Touvron, Matthieu Cord, Alaaeldin El-Nouby, Piotr Bojanowski, Armand Joulin, Gabriel Synnaeve, and Hervé Jégou. Augmenting convolutional networks with attention-based aggregation, 2021. URL https://arxiv.org/abs/2112.13692.

Gabriel Tseng, Ivan Zvonkov, Catherine Lilian Nakalembe, and Hannah Kerner. Cropharvest: A global dataset for crop-type classification. In *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2)*, 2021.

Gabriel Tseng, Anthony Fuller, Marlena Reil, Henry Herzog, Patrick Beukema, Favyen Bastani, James R Green, Evan Shelhamer, Hannah Kerner, and David Rolnick. Galileo: Learning global & local features of many remote sensing modalities. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu (eds.), *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 60280–60300. PMLR, 13–19 Jul 2025. URL https://proceedings.mlr.press/v267/tseng25a.html.

Gabriel Tseng, Yawen Zhang, Favyen Bastani, Henry Herzog, Joseph Redmon, Hadrien Sablon, Piper Wolters, Ando Shah, Patrick Alan Johnson, Christopher Wilhelm, and Patrick Beukema. Olmoearth v1.2: A more efficient family of olmoearth models, 2026. URL https://arxiv.org/abs/2605.20804.

Ana M. Tárano, Gray Tappan, Hunter Blevins, Catlynh Nguyen, Iman Smith, Christopher Atsianzale Wakhanala, Gedeon Muhawenayo, Inbal Becker-Reshef, Karyn Tabor, Amy McNally, and Hannah Kerner. Sudan active cropland reference points (2022-2024), April 2026. URL https://doi.org/10.5281/zenodo.19600105.

Josef Wagner, Shabarinath S. Nair, Sergii Skakun, Erik C. Duncan, Fangjie Li, Oleksandra Olinyk, Françoise Nerry, Jean Rehbinder, and Inbal Becker-Reshef. Monitoring cropland cultivation, abandonment, fallowing and recultivation dynamics with regard to conflict intensity in war-affected ukraine. *Science of Remote Sensing*, 12:100326, 2025. ISSN 2666-0172. doi: https://doi.org/10.1016/j.srs.2025.100326. URL https://www.sciencedirect.com/science/article/pii/S2666017225001324.

Ivan Zvonkov, Gabriel Tseng, Inbal Becker-Reshef, and Hannah Kerner. Cropland mapping using geospatial embeddings, 2025. URL https://arxiv.org/abs/2511.02923.