The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLIX-B3-2026
XXV ISPRS Congress 2026 “From Imagery to Understanding”, Commission III, 4–11 July 2026, Toronto, Canada

Mowing Event Temporal Localization on Dense Satellite Time Series Using Foundational Models

Tilemachos Moumouris[1], Vasileios Tsironis[1], Athena Psalta[1], Konstantinos Karantzalos[1]

[1]NTUA, Remote Sensing Lab, 15772 Zografou Athens, Greece
tmoumouris@mail.ntua.gr
{tsironisbi, psaltaath, karank}@central.ntua.gr

Keywords: Mowing Events, Temporal Localization, Foundational Models, Optical Data

Abstract

Mowing event temporal localization in dense satellite time series is essential for agricultural monitoring and Common Agricultural Policy (CAP) compliance. In this study, we investigate the use of Foundation Models (FM) as frozen encoders which construct useful features for precise temporal mowing event localization. We study the Prithvi-EO-2.0 family of FM, at different encoder sizes, paired with a lightweight MLP localization head, and compare against a 3D CNN baseline. Input timeseries from the Harmonized Landsat Sentinel-2 (HLS) dataset, are comprised of observations across six common spectral bands over the growing season in Central Greece. Moreover, a sliding window approach is used to process the input data adding tunable hyperparameters to the architecture. We further contribute 451 newly annotated parcels with Day of Year(DOY) labels for each event that occurred, extending an existing benchmark for standardized mowing event monitoring evaluation. Our results reveal a connection between event detection and temporal localization. The 3D CNN baseline achieves the highest F1 score of 83.63%, while the Prithvi-EO-2.0 600M variant attains the best Mean Temporal Error (MTE) of 0.83 timesteps, representing a 33% reduction compared to the baseline model. On top, we conduct an ablation study across model sizes, window sizes, and tolerance values which shows that scaling in a certain area of model size yields the largest performance gain. Additionally, window size and tolerance of detection are important for performance across all model sizes. The dataset and code are available at the official repo.

1. Introduction

Grassland management monitoring over large agricultural areas is critical for Common Agricultural Policy (CAP) compliance at both national and European levels. Mowing events in cultivated grasslands serve as a direct indicator of parcel management intensity and their precise monitoring is critical not only for verifying farmer declarations but also for biodiversity conservation, as grassland use intensity is tightly linked to habitat quality and farmland bird abundance (Andreatta et al., 2025). Earth Observation (EO) time series, owing to their non-invasive nature, broad spatial coverage and high revisit frequency, have emerged as the primary operational tool for delivering timely, parcel-level evidence to decision makers without reliance on costly in-situ inspections (KARAKIZI et al., 2024). Existing methods for mowing monitoring, spanning rule-based thresholding, data-driven CNNs with single or multi-modal inputs, and knowledge distillation approaches, have demonstrated strong performance in estimating how many mowing events occur within a season (Schwieder et al., 2022, Kolecka et al., 2018, Lobert et al., 2021, Holtgrave et al., 2023, Miletich et al., 2025, Moumouris et al., 2025). However, they predominantly address event frequency and at best approximate timing at a coarse granularity. The precise temporal localization of each mowing event, i.e. recovering its exact Day of Year (DOY) within the time series, remains largely unaddressed. This distinction is operationally critical since knowing when a mowing event occurred within a narrow temporal window is far more actionable for CAP compliance verification and practice characterization than simply counting events. On the representation learning side, EO Foundation Models (FMs) pre-trained on large-scale multi-sensor archives have substantially advanced geospatial feature extraction across spatial tasks such as classification, segmentation, and change detection (Bommasani et al., 2021, Cong et al., 2022, Xiong et al., 2024, Wang et al., 2025, Danish et al., 2025). Among these, Prithvi-EO-2.0 (Szwarcman et al., 2026) stands apart by integrating dedicated temporal and geolocation embeddings trained on 4.2M global HLS time series, making it inherently suited for reasoning about when within a seasonal cycle an observation occurs. Yet its capability for fine-grained per-event temporal localization, which is a fundamentally short-duration, temporal detection problem, has not been explored. In this work, we address this gap by proposing a frozen Prithvi-EO-2.0 encoder paired with a lightweight MLP localization head and a sliding window processing strategy for precise mowing event temporal localization on dense HLS time series. We validate our approach through systematic experiments and contribute a new annotated benchmark to support future research. Specifically, our contributions are:

- A new annotated dataset of 451 grassland parcels in Central Greece, labeled with precise per-event Day of Year (DOY) mowing timestamps, extending an existing benchmark beyond frequency classification.
- A frozen-encoder framework combining Prithvi-EO-2.0 at four scales (Tiny, 100M, 300M, 600M) with a lightweight MLP head, demonstrating effective temporal event localization under limited supervision.
- A sliding window processing strategy that enables per-event localization throughout the full growing season with tunable temporal resolution.
- Comprehensive experimental analysis revealing a detection–localization trade-off, with Prithvi-EO-2.0 600M version achieving a 33% reduction in Mean Temporal Error over the 3D CNN baseline.

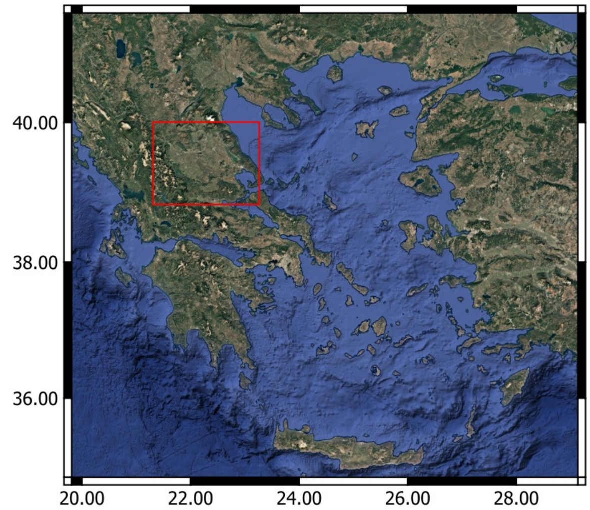

Figure 1. Study area of newly added parcels in Central Thessaly.

2. Related Work

Mowing event detection and localization. Mowing event detection from satellite time series encompasses two complementary sub-tasks: *frequency detection*, where the number of events within a growing season is estimated, and *temporal localization*, where the specific Day of Year (DOY) of each event is predicted. Both tasks rely on the characteristic spectral signature of mowing, namely an abrupt decline in vegetation indices such as NDVI or EVI followed by gradual recovery, a pattern that is only observable at sufficient temporal density and with appropriate cloud screening. The ecological relevance of this monitoring task extends beyond agricultural policy, as precise mowing date estimation has been linked to landscape-scale effects on farmland bird abundance (Andreatta et al., 2025), while grassland spectral parameter models have been shown to be highly sensitive to the timing of satellite acquisitions relative to management events such as cutting (Wallis et al., 2025). Rule-based methods applied to combined Sentinel-2 and Landsat-8 time series have demonstrated scalable mowing frequency mapping at regional and national levels (Schwieder et al., 2022), (Kolecka et al., 2018). Moving toward data-driven localization, 1D CNNs trained on multi-modal optical and SAR features have been shown to recover event timing with competitive temporal precision (Lobert et al., 2021), while the inclusion of weather variables alongside satellite observations has further confirmed that optical data remains the most discriminative source for event detection despite the benefit of multimodal fusion (Holtgrave et al., 2023). A training-free approach combining NDVI drop detection with complementary SWIR increase filtering produced the first open, wall-to-wall mowing event dataset for Austria, demonstrating that high-quality frequency monitoring is attainable without labeled training data (Miletich et al., 2025). A benchmark of over 1,600 annotated parcels across three regions of Greece, built on HLS time series and covering up to four mowing events per season, established that deep learning architectures combined with knowledge distillation on pseudo-labeled data can yield strong frequency classification results even under data scarcity (Moumouris et al., 2025). The present work builds on this foundation by expanding the benchmark with newly annotated parcels carrying per-event DOY labels, and shifts the objective from frequency classification to precise per-event temporal localization, a task that demands not only detecting that a mowing event occurred but recovering its exact position within the time series.

Foundational models for remote sensing. Foundation Models (FMs) are large-scale architectures pre-trained on vast, diverse datasets to serve as general-purpose backbones transferable with minimal task-specific supervision to a wide range of downstream tasks such as classification, segmentation, object detection and change detection (Bommasani et al., 2021). The remote sensing community has embraced this paradigm to overcome the persistent bottleneck of labeled data scarcity, leading to a growing ecosystem of EO-specific FMs. Early efforts such as SatMAE (Cong et al., 2022) demonstrated that extending the Masked Autoencoder (MAE) (He et al., 2022) framework with explicit temporal and spectral positional encodings yields strong transfer learning gains on land cover classification and segmentation tasks, establishing temporal embedding as a viable strategy for satellite image pre-training. DOFA (Xiong et al., 2024) addressed sensor heterogeneity through a neuroplasticity-inspired wavelength-conditioned hypernetwork capable of ingesting SAR, multispectral and hyperspectral data under a unified architecture. Copernicus-FM (Wang et al., 2025) introduced flexible metadata encoding and cross-mission multi-modal pre-training across all Sentinel satellites. TerraFM (Danish et al., 2025) demonstrated the scalability of this approach, achieving state-of-the-art results on GEO-Bench(Lacoste et al., 2023) through joint Sentinel-1/2 training. While these models have collectively advanced the state of EO representation learning, their focus remains predominantly on spatial tasks, and temporal ordering is typically treated as an auxiliary, rather than structural, component of the architecture. Prithvi-EO (NASA-IBM) (Jakubik et al., 2023), by contrast, was designed from the ground up for multi-temporal EO analysis. It was pre-trained on 4.2M global time series from the HLS archive and its architecture integrates dedicated temporal and geolocation embeddings that allow the model to reason about when an observation occurs within the seasonal cycle. Its successor, Prithvi-EO-2.0 (Szwarcman et al., 2026), further scaled this design to 300M and 600M parameter variants, extending multi-temporal pre-training with improved geospatial positional encodings and demonstrating stronger generalization across a broader range of EO downstream tasks. Prithvi's native temporal awareness makes it the most suitable FM backbone among current alternatives for our task, since mowing event localization requires detecting subtle, short-duration reflectance transitions within dense satellite time series, a problem that is fundamentally temporal in nature and benefits directly from a model pre-trained to distinguish phenological stages across time.

3. Dataset

3.1 Study Area

This study focused on Central Thessaly (Figure 1), a region characterized by high-intensity agricultural activity. To ensure consistency with existing benchmarks (Moumouris et al., 2025), NASA's Harmonized LandSat Sentinel (HLS) Dataset (Claverie et al., 2018) was utilized as the primary data source. The temporal scope of the data covers the cultivation period from early April until late November for year 2021. To obtain high-density time series with minimal atmospheric interference, a maximum cloud cover threshold of 20% was enforced during data preparation, and default quality filtering was applied according to the official HLS Dataset guidelines. The resulting time series consists of 27 timesteps at a spatial resolution of 30m, incorporating spectral bands from both sensors.

3.2 New Dataset

Building upon the dataset available in (Moumouris et al., 2025), an additional 451 parcels were collected to create a specialized fine-tuning set. This new subset provides high-resolution labels, including the specific Day of Year (DOY) for each mowing event and the total count of events per parcel. The annotation process was performed according to the following steps and is also described in Figure 2:

1. **Enhanced Vegetation Index (EVI) calculation:** EVI was calculated for each timestep to identify sudden drops in the index, which typically signify a mowing event.
2. **Optical (RGB) and EVI time series construction:** The synergy of both data sources was utilized to verify visual changes against vegetation index trends for more robust results.
3. **Parcel annotation:** Each parcel in the Region of Interest (ROI) was assigned two distinct labels: the total number of events and the specific DOY for each occurrence.

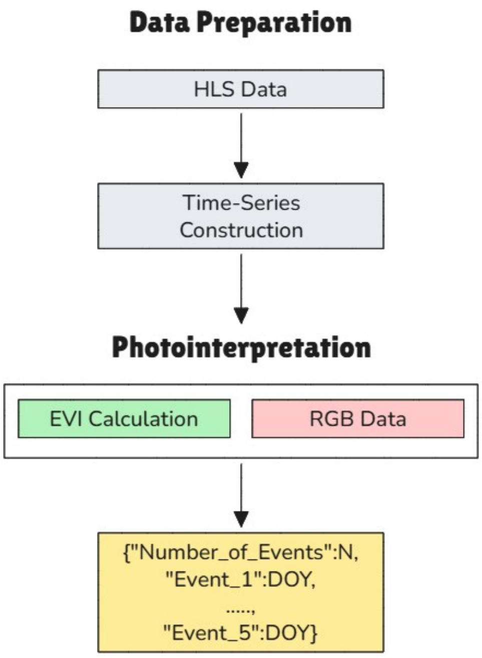

Figure 2. Flowchart describing the annotation process of the newly added part of the fine-tuning dataset.

The small size of the parcels posed a significant challenge during selection, requiring careful identification of parcels representative of the broader agricultural landscape in the area. Notably, quality filtering resulted in missing values that were intentionally left un-imputed to facilitate future research into missing value reconstruction methods. The final fine-tuning set contains five classes based on the total number of events, ranging from one to five DOYs per parcel; only parcels with at least one confirmed mowing event were included.

As shown in Table 1, the distribution is dominated by parcels with 3 or 4 events, which account for over 70% of the dataset. Before training, a data preprocessing step is applied to the dataset to produce temporal windows of different sizes. This sliding windows approach is described in detail in the next section and aims to localize the events present in specified time windows, enabling event detection throughout the cultivation season. Moreover, in this dataset, only the six common channels between Landsat-8 and Sentinel-2 were selected apart from the Enhanced Vegetation Index (EVI), in order to better capture parcel dynamics.

Events per parcel	Number of parcels	Percentage (%)
1 event	27	6.0
2 events	58	12.9
3 events	155	34.4
4 events	169	37.5
5 events	42	9.3
Total	451	100.0

Table 1. Distribution of annotated mowing-event counts per parcel (fine-tuning set).

4. Methodology

4.1 Sliding Temporal Windows

Time series data was treated as individual windows as input to the model, implementing a sliding window approach, thus introducing two hyperparameters: window size W and stride S. Window start indices are generated as $\{0, S, 2S, ...\}$ over the full time series of length T, with a final window always appended at position $T - W$ to guarantee complete coverage regardless of divisibility. During fine-tuning, two different window sizes were selected,5 and 3 timesteps, with the model processing all frames at once to better capture any underlying relations. Furthermore, a tolerance offset factor was introduced during training and evaluation to allow minor temporal prediction errors to be treated as correct, when they occur within this tolerance. For this study, taking under consideration the selected window sizes, a tolerance of 1 or 2 timesteps was studied. An example of the sliding window approach is given in Figure 3.

Figure 3. Sliding window strategy and time series details. The depicted example shows a time series and the sliding window strategy followed in our study.

4.2 Encoder Model Selection

In this study we utilized all available common bands between Sentinel-2 and LandSat in the HLS dataset, namely: BLUE, GREEN, RED, NIR, SWIR 1, SWIR 2. The model under study was Prithvi-EO-2.0 as a frozen encoder, giving the opportunity to study the use of this kind of models in limited data cases. No additional temporal or location embeddings were utilized as only the output features of the frozen encoder were used in each model version. This model family, was pretrained using the HLS data, thus making it an excellent use case for our dataset.

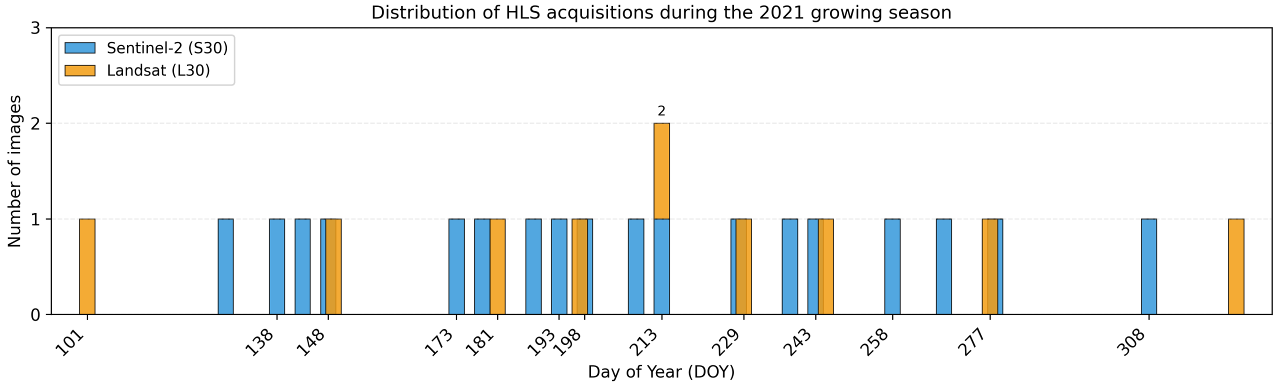

Figure 4. Distribution of observations during the growing season.

4.3 Localization Head

A modular classification head maps the encoded feature sequence (B, T, D) to per-timestep logits (B, T). The head architecture was chosen as an MLP, with a hidden dimension of 256 and a dropout rate of 0.1 as defaults before any fine-tuning. A temporal mask (B, T) is passed to the head to suppress contributions from invalid or missing timesteps during both training and inference.

4.4 Postprocessing

Raw per-timestep logits are passed through a sigmoid activation and a fixed threshold of $\tau = 0.5$ then converts outputs to binary predictions. Prior to encoding, an optional spatial mask (B, T, H, W) is applied to eliminate invalid pixels, such as those contaminated by clouds or missing data, ensuring that non-valid observations do not propagate into the feature representations.

5. Results

5.1 Implementation Details

We implement all models using PyTorch and TerraTorch frameworks and train them on a single GPU. For all Prithvi-EO-2.0 variants, we utilize the pretrained weights and keep the encoder frozen, fine-tuning only the localization head on our newly annotated dataset of 451 parcels. All models are optimized using Adam with a learning rate of 1×10^{-3} and trained for up to 100 epochs with early stopping based on validation MTE. The 3D CNN baseline follows the same training protocol for a fair comparison.

We evaluate two model families: (i) a lightweight 3D CNN baseline and (ii) the Prithvi-EO-2.0 foundation model at four scales (Tiny, 100M, 300M, 600M). Also, we report two primary metrics, namely F1-score for event detection (binary presence/absence per window) and Mean Temporal Error (MTE), defined as the mean absolute difference in timesteps between predicted and ground-truth event positions, which serves as the primary cross-model localization metric.

5.2 Experimental Results

3D CNN Baseline. The 3D CNN baseline achieves the highest detection performance, reaching 83.63% F1 with tolerance=2 and W=3. Its MTE remains between 1.24 and 1.60 timesteps, corresponding to a temporal offset of approximately 10 to 14 days. This behavior is consistent with the limited temporal receptive field of a shallow architecture, which captures coarse spectral changes but struggles to resolve the precise onset of vegetation recovery patterns. Furthermore, since the baseline is trained from scratch on only 451 parcels with a comparatively larger number of trainable parameters than the frozen-head Prithvi variants, its higher F1 may partly reflect sensitivity to the training distribution rather than robust generalization, a risk the frozen encoder design is specifically intended to mitigate.

Prithvi-EO-2.0. Fine-tuning the frozen Prithvi-EO-2.0 encoder yields moderately lower F1 scores (77.15% for the 100M variant at best under iso-tolerance conditions) but substantially improved temporal localization. The 600M variant achieves the best MTE of 0.83 timesteps ($\approx$ 7 days based on the distribution of observations, Figure 4), representing a 33% reduction over the best baseline MTE of 1.24. The 300M model attains a competitive MTE of 0.86 timesteps, confirming that larger pretrained encoders consistently benefit temporal precision. This improvement is attributable to the richer spatio-temporal representations learned during large-scale HLS pre-training, which enable the model to identify subtle but diagnostic spectral transitions more precisely than a shallow CNN trained from scratch.

Detection and Localization Trade-Off. A consistent architectural trade-off emerges across all configurations. The 3D CNN baseline prioritizes event detection at the cost of temporal precision, while Prithvi-EO-2.0 variants achieve tighter localization with moderately lower detection rates. From an operational perspective, this distinction is consequential; for CAP compliance monitoring, recovering the precise timing of a management event within a narrow temporal window carries more evidentiary value than binary occurrence confirmation alone.

5.3 Ablation Studies

To disentangle the contributions of model capacity, window size, and tolerance, we conduct a systematic ablation across all Prithvi-EO-2.0 variants. Results are reported in Table 2.

Effect of Model Size. Scaling from Tiny to 100M yields the largest single improvement: F1 increases from 71.08% to 77.15% adding a 6.07%, and MTE drops from 2.27 to 0.98, a 57% decrease. The Tiny model's poor performance suggests that its limited embedding dimension is insufficient to encode the temporal dynamics of mowing events from only 3

https://doi.org/10.5194/isprs-archives-XLIX-B3-2026-1009-2026

Model	F1 (%)	Window Size	Tolerance	MTE
3D CNN (baseline)	79.37	3	1	1.60
3D CNN (baseline)	83.63	3	2	1.24
Prithvi-EO-2.0 (tiny)	71.08	3	1	2.27
Prithvi-EO-2.0 (tiny)	42.87	5	2	1.61
Prithvi-EO-2.0 (tiny)	41.52	5	1	1.59
Prithvi-EO-2.0 (100M)	77.15	3	1	0.98
Prithvi-EO-2.0 (100M)	54.18	5	1	1.59
Prithvi-EO-2.0 (100M)	56.81	5	2	1.60
Prithvi-EO-2.0 (300M)	57.12	5	1	1.16
Prithvi-EO-2.0 (300M)	71.56	3	1	0.86
Prithvi-EO-2.0 (300M)	55.98	5	2	1.17
Prithvi-EO-2.0 (600M)	56.51	5	1	1.50
Prithvi-EO-2.0 (600M)	51.88	5	2	1.50
Prithvi-EO-2.0 (600M)	75.00	3	1	**0.83**

Table 2. Model performance across architectures. All models’ performance and related hyper-parameters are reported for comparison.

to 5 frames. Further scaling to 300M and 600M improves MTE marginally (0.86 and 0.83 respectively) but does not increase F1 beyond the 100M level. This pattern of diminishing returns is consistent with the small fine-tuning set of 451 parcels. Firstly, larger encoders produce more precise temporal features, but on the other hand, the lightweight MLP head cannot fully exploit the additional capacity for detection. Finally, this suggests that scaling the classification head or incorporating temporal attention mechanisms could unlock further gains for larger encoders.

Effect of Window Size. Across all Prithvi variants, $W=3$ consistently outperforms $W=5$ in both F1 and MTE. The degradation with $W=5$ is more obvious for the 300M model where F1 drops from 71.56% to 57.12% and MTE increases from 0.86 to 1.16 timesteps. A factor that is likely to explain this behavior is the large window that dilutes the mowing pattern. With a mean revisit interval of approximately 8 days, a 5-frame window spans for 40 days, during which multiple mowing events may occur, making it difficult for the crop to recover on time, so the phenological pattern repeats.

Effect of Tolerance. Tolerance has a more modest effect than window size. Critically, increasing tolerance inflates F1 without improving actual localization precision. Models evaluated at different tolerance values must not be compared on F1 alone; MTE remains the only reliable cross-condition localization metric, capturing temporal offset differences that F1 with an adjusted threshold cannot distinguish.

6. Conclusions & Future Work

This work addresses the largely unexplored problem of precise mowing event temporal localization in dense satellite time series, shifting the objective from frequency classification to per-event Day of Year prediction. In this study we explore the use of Foundation Models for mowing event localization, utilizing them as frozen encoders, and offer a fair comparison between the Prithvi-EO-2.0 model family and a 3D CNN architecture in the temporal localization task. To support this objective, a fine-tuning dataset is constructed leveraging all available common bands between Sentinel-2 and Landsat missions, extending the benchmark we previously proposed with precise per-event DOY annotations for 451 newly labeled parcels in Central Greece, which is the first publicly available resource of this kind for the Mediterranean region, to our knowledge.

Our results suggest that a clear trade-off is present between mowing event detection, treated as a binary task and temporal localization. The 3D CNN baseline achieves strong detection performance but localizes events with a temporal offset of 10–14 days, a limitation consistent with its shallow architecture and lack of pre-trained temporal representations. In contrast, the 600M variant achieves the best temporal precision with an MTE of 0.83 timesteps, which translates to approximately 7 days based on our dataset characteristics. This reduction of 33% gives the opportunity to better monitor parcels, especially small-sized ones that are the majority in Greece. The ablation further reveals that the most critical scaling gain occurs between Tiny and 100M, while W=3 consistently outperforms W=5 by preserving the short-duration spectral signature of mowing events within the sliding window.

Due to the findings in this study there are several future directions that emerge for performance improvement. The small fine-tuning set of 451 parcels is the primary bottleneck constraining both detection F1 and the ability of larger encoders to fully express their representational capacity. The amount of data could be increased not only in temporal coverage but also by adding more areas in Greece and possibly other countries of the Mediterranean region. The intentional absence of gap imputation in our dataset opens a direct research avenue into missing value reconstruction and in future studies a robust gap handling algorithm will be of great importance, especially in temporal localization tasks. On the architecture side, replacing the lightweight MLP head with a temporal attention mechanism could better exploit the richer feature space of the 300M and 600M encoders. Also, different modalities could be added to the existing dataset, such as Sentinel-1 or temperature data, as they could lead to better results, particularly in cloud-affected periods where optical data density is reduced. Due to the strong generalization of these models, more frozen encoders are to be studied, leading to faster development of monitoring solutions due to the pretrained nature of the backbone.

References

Andreatta, D., Bazzi, G., Nardelli, R., Siddi, L., Cecere, J. G., Chamberlain, D., Morganti, M., Rubolini, D., Assandri, G., 2025. Remote sensing reveals scale-specific effects of forage crop mowing and landscape structure on a declining farmland bird. *Journal of Applied Ecology*, 62(3), 502–515.

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E. et al., 2021. On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*.

Claverie, M., Ju, J., Masek, J. G., Dungan, J. L., Vermote, E. F., Roger, J.-C., Skakun, S. V., Justice, C., 2018. The Harmonized Landsat and Sentinel-2 surface reflectance data set. *Remote Sensing of Environment*, 219, 145-161.

Cong, Y., Khanna, S., Meng, C., Liu, P., Rozi, E., He, Y., Burke, M., Lobell, D., Ermon, S., 2022. Satmae: Pre-training transformers for temporal and multi-spectral satellite imagery. *Advances in Neural Information Processing Systems*, 35, 197–211.

Danish, M. S., Munir, M. A., Shah, S. R. A., Khan, M. H., Anwer, R. M., Laaksonen, J., Khan, F. S., Khan, S., 2025. Terrafm: A scalable foundation model for unified multisensor earth observation.

He, K., Chen, X., Xie, S., Li, Y., Dollár, P., Girshick, R., 2022. Masked autoencoders are scalable vision learners. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 16000–16009.

Holtgrave, A.-K., Lobert, F., Erasmi, S., Röder, N., Kleinschmit, B., 2023. Grassland mowing event detection using combined optical, SAR, and weather time series. *Remote Sensing of Environment*, 295, 113680.

Jakubik, J., Roy, S., Phillips, C. E., Fraccaro, P., Godwin, D., Zadrozny, B., Szwarcman, D., Gomes, C., Nyirjesy, G., Edwards, B., Kimura, D., Simumba, N., Chu, L., Mukkavilli, S. K., Lambhate, D., Das, K., Bangalore, R., Oliveira, D., Muszynski, M., Ankur, K., Ramasubramanian, M., Gurung, I., Khallaghi, S., Hanxi, Li, Cecil, M., Ahmadi, M., Kordi, F., Alemohammad, H., Maskey, M., Ganti, R., Weldemariam, K., Ramachandran, R., 2023. Foundation models for generalist geospatial artificial intelligence.

KARAKIZI, C., KARANTZALOS, K., KANDYLAKIS, Z., 2024. Exploiting Multitemporal Multispectral High-resolution Satellite Data toward Annual Land Cover and Crop Type Mapping: A Case Study in Greece. *Multitemporal Earth Observation Image Analysis: Remote Sensing Image Sequences*, 81–122.

Kolecka, N., Ginzler, C., Pazur, R., Price, B., Verburg, P. H., 2018. Regional scale mapping of grassland mowing frequency with Sentinel-2 time series. *Remote Sensing*, 10(8), 1221.

Lacoste, A., Lehmann, N., Rodriguez, P., Sherwin, E., Kerner, H., Lütjens, B., Irvin, J., Dao, D., Alemohammad, H., Drouin, A. et al., 2023. Geo-bench: Toward foundation models for earth monitoring. *Advances in Neural Information Processing Systems*, 36, 51080–51093.

Lobert, F., Holtgrave, A.-K., Schwieder, M., Pause, M., Vogt, J., Gocht, A., Erasmi, S., 2021. Mowing event detection in permanent grasslands: Systematic evaluation of input features from Sentinel-1, Sentinel-2, and Landsat 8 time series. *Remote Sensing of Environment*, 267, 112751.

Miletich, P., Kirchmair, M., Deutscher, J. G., Schippl, A., Hirschmugl, M., 2025. Open and Free Sentinel-2 Mowing Event Data for Austria. *Remote Sensing*, 17(10), 1769.

Moumouris, T., Tsironis, V., Psalta, A., Karantzalos, K., 2025. Large Scale Mowing Event Detection on Dense Time Series Data Using Deep Learning Methods and Knowledge Distillation. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVIII-M-7-2025, 43–48.

Schwieder, M., Wesemeyer, M., Frantz, D., Pfoch, K., Erasmi, S., Pickert, J., Nendel, C., Hostert, P., 2022. Mapping grassland mowing events across Germany based on combined Sentinel-2 and Landsat 8 time series. *Remote Sensing of Environment*, 269, 112795.

Szwarcman, D., Roy, S., Fraccaro, P., orsteinn Elí Gíslason, Blumenstiel, B., Ghosal, R., de Oliveira, P. H., de Sousa Almeida, J. L., Sedona, R., Kang, Y., Chakraborty, S., Wang, S., Gomes, C., Kumar, A., Truong, M., Godwin, D., Lee, H., Hsu, C.-Y., Lal, R., Asanjan, A. A., Mujeci, B., Shidham, D., Keenan, T., Arevalo, P., Li, W., Alemohammad, H., Olofsson, P., Hain, C., Kennedy, R., Zadrozny, B., Bell, D., Cavallaro, G., Watson, C., Maskey, M., Ramachandran, R., Moreno, J. B., 2026. Prithvi-eo-2.0: A versatile multi-temporal foundation model for earth observation applications.

Wallis, C. I., Holtgrave, A.-K., Prati, D., Förster, M., Kleinschmit, B., 2025. Modeling grassland parameters with hyperspectral satellite data: Comparison of sensors, acquisition times and spectral transformations. *International Journal of Applied Earth Observation and Geoinformation*, 144, 104857.

Wang, Y., Xiong, Z., Liu, C., Stewart, A. J., Dujardin, T., Bountos, N. I., Zavras, A., Gerken, F., Papoutsis, I., Leal-Taixé, L., Zhu, X. X., 2025. Towards a unified copernicus foundation model for earth vision.

Xiong, Z., Wang, Y., Zhang, F., Stewart, A. J., Hanna, J., Borth, D., Papoutsis, I., Saux, B. L., Camps-Valls, G., Zhu, X. X., 2024. Neural plasticity-inspired multimodal foundation model for earth observation. *arXiv preprint arXiv:2403.15356*.