JOJ Horticulture & Arboriculture
ISSN: 2641-8215

Juniper Publishers
key to the Researchers

Review Article
Volume 6 Issue 4 - November 2026
DOI: 10.19080/JOJHA.2026.06.555695

JOJ Hortic Arboric
Copyright © All rights are reserved by Zeenat Khan

Geospatial Foundation Models for Intelligent Crop Monitoring: AI-Driven Multispectral Analysis, Yield Prediction, and Precision Agriculture

Wassaf Ahmad[1], Daud Khan[2] and Zeenat Khan[3]*

[1]*Institute of Geo-Information & Earth Observation, PMAS-Arid Agriculture University, Rawalpindi, Pakistan.*

[2]*Resident Civil Engineer, Al-Mussawir Engineers, Rawalpindi, Pakistan*

[3]*Research Analyst, Al-Mussawir Engineers, Rawalpindi, Pakistan*

Submission: October 10, 2026; **Published:** November 23, 2026

***Corresponding author:** Zeenat Khan, Research Analyst, Al-Mussawir Engineers, Rawalpindi, Pakistan

Abstract

Multispectral remote sensing has become a central source of spatially explicit information for crop mapping, phenology assessment, stress monitoring, and yield estimation. Many agricultural workflows still depend on hand-crafted vegetation indices or task-specific machine-learning models whose performance can deteriorate when crop calendars, management practices, field sizes, sensors, or climatic conditions change. Recent advances in geospatial foundation models (GFMs) introduce a complementary paradigm: large models pretrained on broad Earth-observation archives can learn reusable representations before task-specific adaptation. This review critically examines the emerging role of GFMs in multispectral crop monitoring and yield enhancement.

The review (i) synthesizes the spectral and sensor foundations of crop observation, (ii) organise GFM approaches by pretraining objective, input handling, and output representation, (iii) compare representative models and agricultural evaluation resources, (iv) examine a complementary paradigm: large models pretrained on broad Earth-observation archives can learn reusable representations before task-specific adaptation. This review critically examines the emerging role of GFMs in multispectral crop monitoring and yield enhancement. The review (i) synthesise the spectral and sensor foundations of crop observation, (ii) organise GFM approaches by pretraining objective, input handling, and output representation, (iii) compare representative models and agricultural evaluation resources, (iv) examine applications in crop mapping, phenology and stress monitoring, yield estimation, and UAV- satellite integration, and (v) analyse the gap between predictive performance and agronomic action.

The evidence indicates that GFMs can reduce dependence on labelled data and facilitate transfer, but the corresponding agricultural value remains task- and region-dependent. Strong local performance can coexist with limited spatial transferability, temporal sensitivity, and interpretability, while multimodal models provide new opportunities for earlier stress detection under climate extremes. The analysis therefore identifies priorities for agriculture-specific pretraining, spatially honest validation, multimodal and scale-aware learning, physics-informed integration, uncertainty quantification, and farm-level outcome evaluation.

Keywords: Geospatial Foundation Models; Multispectral Imagery; Crop Monitoring; Yield Prediction; Self-Supervised Learning; Sentinel-2; Precision Agriculture; Remote Sensing

Abbreviations: GFMs: Geospatial Foundation Models, UVA: Unmanned Aerial Vehicle

Introduction

Agricultural production must increase or stabilise productivity while responding to climate variability, water constraints, soil degradation, and the need to reduce unnecessary inputs. Achieving these objectives requires timely, spatially explicit information on crop distribution, developmental stage, canopy condition, and expected productivity [1]. Satellite and unmanned aerial vehicle (UAV) multispectral observations provide complementary spatial and temporal information and are increasingly used for crop mapping, crop-condition assessment, phenology monitoring, and yield estimation [2]. These considerations are particularly important when models are transferred between production systems with different cultivars, planting dates, irrigation regimes,

and background soil conditions [3]. Consequently, a scientifically useful monitoring framework should treat the imagery as a sequence of physically meaningful observations rather than as a collection of independent images. [4] Recent agricultural remote-sensing reviews emphasise the growing role of multispectral imagery, data fusion, and machine learning in operational crop monitoring, while Sentinel-2 has become a particularly important source because of its spatial, spectral, and revisit characteristics. The principal challenge is therefore no longer simply access to imagery, but the reliable conversion of heterogeneous observations into estimates that remain useful across locations, seasons, and management systems.

Conventional workflows commonly derive vegetation indices and fit empirical or machine-learning relationships to agronomic variables [5]. Deep learning can learn richer spatial and temporal representations, and recent yield-prediction literature increasingly combines optical imagery with weather, radar, and other covariates [6]. Nevertheless, performance can remain sensitive to training-region composition, label quality, temporal coverage, and distribution shift [7]. Recent agricultural remote-sensing studies emphasize that crop observations are inherently spatiotemporal, because the same field can exhibit substantially different spectral responses as canopy structure and phenological stage change [8]. The value of multispectral information therefore depends not only on spectral discrimination but also on acquisition timing, atmospheric quality, field geometry, and the availability of observations close to agronomic decision points [9]. These limitations are especially consequential where field observations are sparse or where smallholder production systems differ substantially from the regions represented in training data. Foundation models provide a different learning strategy by separating broad representation learning from downstream task adaptation.

In Earth observation, recent models such as SatMAE, Prithvi, CROMA, DOFA, Clay, Galileo, and AlphaEarth Foundations demonstrate different ways of exploiting large unlabelled or weakly labelled archives through masked reconstruction, contrastive learning, multimodal fusion, wavelength conditioning, or embedding-field generation [10]. The potential agricultural advantage is label efficiency: a pretrained representation may provide useful structure before only a small amount of crop-specific supervision is introduced [11]. Recent foundation-model research increasingly addresses these constraints through temporal embeddings, wavelength conditioning, multimodal alignment, and scale-aware encoders [12]. The agricultural question is consequently whether these architectural choices improve transfer under realistic label scarcity and distribution shift rather than whether these models simply increase benchmark accuracy [13]. The principal advantage of this paradigm is not model size by itself but the possibility of learning reusable representations from observations for which dense agricultural labels are unavailable [14]. Recent surveys also show, however, that remote-sensing foundation models remain constrained by sensor heterogeneity, spatial and temporal distribution shifts, computational cost, and inconsistent evaluation protocols.

This review addresses that gap by retaining a crop-centred perspective rather than treating agriculture as one downstream benchmark among many. Its contributions are (1): a concise account of the spectral and sensor basis of multispectral crop observation (2); a taxonomy and comparison of twelve GFMs relevant to agricultural applications (3); a critical analysis of crop mapping, phenology and stress monitoring, yield estimation, and UAV–satellite scale bridging (4); three representative case studies (5); an assessment of agricultural benchmarks and evaluation practice (6); and an analysis of the transition from model predictions to agronomic decisions [15]. Self-supervised learning is particularly compatible with Earth observation because satellite archives contain enormous quantities of imagery while expert-labelled agricultural variables remain comparatively scarce [16]. Nevertheless, the learned representation reflects the geographic, spectral, temporal, and preprocessing characteristics of the pretraining corpus, so apparent generality must be demonstrated empirically [17]. The scope is primarily optical multispectral data, while SAR, weather, terrain, and other modalities are discussed where these models are integrated with optical observations.

Review Methodology

The review follows a structured narrative design informed by PRISMA 2020 principles (Table 1). Because the GFM literature is rapidly evolving and includes peer-reviewed articles, conference papers, benchmark reports, and influential preprints, the review does not attempt a quantitative meta-analysis [18]. The search covered Scopus, Web of Science Core Collection, IEEE Xplore, ScienceDirect, arXiv, and Google Scholar, with forward and backward citation chasing [19]. Records were screened by title and abstract and then by full text [20]. Sources were retained when these models (a) introduced or evaluated pretrained Earth-observation models relevant to multispectral or multimodal inputs, or (b) provided substantive evidence on crop mapping, crop monitoring, stress, phenology, or yield estimation [21]. Preprints were retained when these models described influential models or resources for which the technical contribution was directly relevant to the review; the corresponding publication status is indicated in the reference list.

Multispectral Sensing for Crop Monitoring

Spectral basis of crop condition

Green vegetation exhibits a structured spectral response arising from pigment absorption, leaf and canopy scattering, water absorption, and canopy architecture. Chlorophyll produces strong absorption in the blue and red regions and a local reflectance maximum in the green region [22]. Leaf and canopy structure strongly increase NIR reflectance, producing the characteristic red-edge transition between the red and NIR regions, while SWIR wavelengths contain information related to canopy water and dry matter [23]. Radiative-transfer modelling can further connect

How to cite this article: Wassaf A, Zeenat K. Geospatial Foundation Models for Intelligent Crop Monitoring: AI-Driven Multispectral Analysis, Yield Prediction, and Precision Agriculture. JOJ Horticulture & Arboriculture, 6(4). 555695. DOI: 10.19080/JOJHA.2026.06.555695

latent representations to physically interpretable vegetation variables and can provide synthetic examples when field observations are limited [24]. Accordingly, spectral information should be evaluated together with acquisition geometry, atmospheric correction, crop phenology, and background effects rather than interpreted as an isolated predictor [25]. This physical grounding is important if foundation-model embeddings are expected to support agronomic inference rather than only statistical classification [26]. These properties explain the value of red-edge and SWIR measurements for crop-condition assessment and for estimating biophysical and biochemical variables. Crop stress modifies these spectral relationships, although the magnitude and direction of the response depend on crop type, growth stage, canopy structure, background soil, illumination, and stress severity. Water limitation can alter NIR and SWIR responses through changes in leaf water content and canopy structure; reductions in chlorophyll can modify visible and red-edge reflectance; and disease or pest damage can produce more complex spatial and spectral signatures.

Table 1: Search and selection strategy.

Element	Specification
Databases	Scopus, Web of Science Core Collection, IEEE Xplore, ScienceDirect, arrive (preprints), Google Scholar (forward and backward citation chasing)
Period	January 2017 to September 2026
Core query	("foundation model" OR "self-supervised" OR pretrain∗ OR "masked autoencoder") AND (multispectral OR "Sentinel-2" OR Landsat OR HLS) AND (crop∗ OR agriculture∗ OR yield OR phenology∗)
Inclusion	Pretrained EO models using multispectral input; crop mapping, stress, phenology or yield studies with multispectral data; benchmark and dataset papers; review articles
Exclusion	RGB-only web imagery models; studies without any agricultural or multispectral relevance; non-English full texts; duplicate reports
Synthesis	Narrative synthesis organized by model family, application and evaluation practice; qualitative appraisal of evidence strength

[27]. Radiative-transfer models such as PROSPECT/SAIL provide a physically interpretable link between canopy properties and reflectance and can therefore support simulation, inversion, and physics-informed learning [28]. The physical interpretation of these bands provides an important constraint on data-driven learning because model representations should ultimately respond to measurable changes in pigments, canopy structure, water content, and dry matter (Figure 1) [29]. Recent reviews show that Sentinel-2 red-edge information is particularly useful for crop biophysical and biochemical retrieval, although relationships developed in one environment may not transfer directly to another [30]. The combination of red-edge and SWIR bands available from Sentinel-2 is consequently useful for separating several crop-condition signals that cannot be represented adequately by RGB imagery alone.

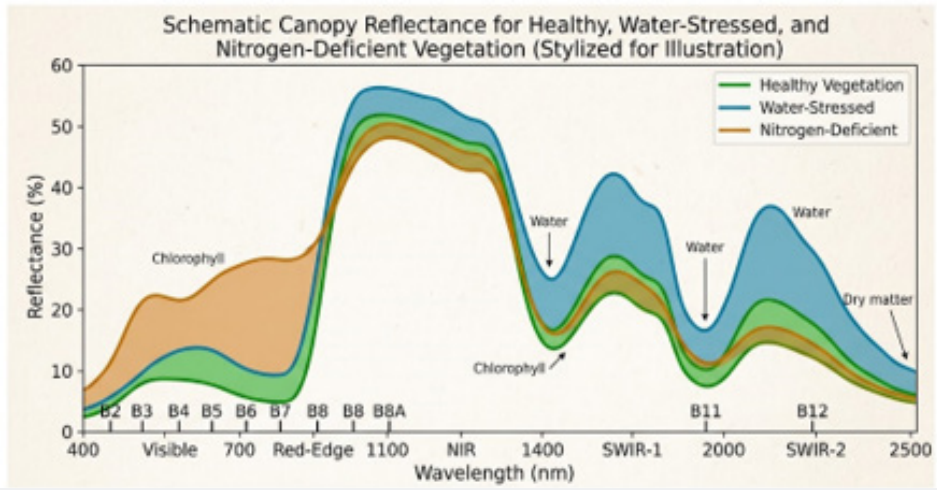


Figure 1: Schematic canopy reflectance for healthy, water-stressed and nitrogen-deficient vegetation, with approximate Sentinel-2 band centers marked. Curves are stylized for illustration and are not measured spectra.

How to cite this article: Wassaf A, Zeenat K. Geospatial Foundation Models for Intelligent Crop Monitoring: AI-Driven Multispectral Analysis, Yield Prediction, and Precision Agriculture. JOJ Horticulture & Arboriculture, 6(4). 555695. DOI: 10.19080/JOJHA.2026.06.555695

Sensing platforms

Multispectral platforms differ in spatial resolution, revisit frequency, spectral configuration, radiometric characteristics, coverage, and cost. Sentinel-2 provides 13 spectral bands at 10, 20, and 60 m spatial resolutions, with a nominal five-day revisit for the two-satellite constellation [31]. Landsat 8 and 9 provide a longer 30 m record, while the Harmonized Landsat and Sentinel-2 (HLS) product improves temporal density and cross-sensor consistency at 30 m (Table 2) [32]. The current HLS Version 2.0 record provides near-global coverage and incorporates improved atmospheric correction, cloud masking, bidirectional-reflectance normalization, and bandpass harmonization [33].

Table 2: Representative multispectral platforms used in crop monitoring.

Platform	Spatial resolution	Revisit	Bands (indicative)	Typical crop use
Sentinel-2 MSI	10 / 20 / 60 m	~5 days (two satellites)	13, incl. 3 red-edge, 2 SWIR	Field-scale mapping, phenology, stress, yield [33]
Landsat 8/9 OLI	30 m	~8 days combined	9 (optical/SWIR)	Long time series, regional yield, historical baselines
HLS (Landsat + Sentinel-2)	30 m	~2-3 days	Harmonized optical/SWIR	Dense time series for phenology and yield [33]
Commercial small sats	~3 m	Near-daily	4-8 bands	Small-field monitoring; license-restricted
MODIS	250-1000 m	Daily	Coarse multispectral	Regional and national yield, drought monitoring
UAV multispectral camera	1-10 cm	On demand	Typically, 4-6 bands	Plot-scale phenotyping, stress scouting, calibration of satellite products

Commercial small-satellite constellations can provide meter-scale observations at high frequency, whereas UAV multispectral systems provide centimeter-scale observations on demand but over much smaller areas and with sensor-specific calibration requirements [34]. This heterogeneity is a fundamental consideration for GFMs because agricultural deployment rarely involves a single sensor or spatial scale. Sensitive indices provide complementary sensitivity under dense canopies, soil-background conditions, chlorophyll variation, or water stress [35]. Recent agricultural reviews show that indices remain valuable for crop discrimination and monitoring because these models are computationally inexpensive, physically interpretable, and easy to reproduce (Table 3) [36]. The corresponding limitation is that these models are predetermined transformations: these models do not explicitly learn crop-, stage-, region-, or sensor-specific representations, and the corresponding relationships with yield or stress can vary across environments [37]. The complementary use of interpretable indices and learned embeddings may provide a more defensible pathway for operational crop monitoring [38]. Vegetation indices remain valuable because the corresponding mathematical structure is transparent and the corresponding relationships with canopy properties can often be inspected directly by agronomists [39]. For this reason, indices should remain important baselines even when GFMs are used as learned feature extractors.

Vegetation indices

Vegetation indices reduce multispectral observations to interpretable combinations of spectral bands that emphasize particular canopy properties. NDVI remains a widely used indicator of vegetation greenness, while EVI, SAVI, GNDVI, red-edge indices such as NDRE, and water.

Evolution of analytical approaches

Analytical approaches have progressed from fixed spectral rules toward increasingly data-driven representation learning (Figure 2). Index-and-threshold methods provide simple and interpretable decision rules [40]. Classical machine-learning methods such as random forests, support-vector machines, and gradient boosting can combine multiple bands, indices, dates, and auxiliary variables, but the corresponding performance often depends on feature engineering and training-data representativeness. [41] Deep-learning approaches learn spatial and temporal features directly and have been applied to in-season crop mapping and yield prediction using multispectral time series and multimodal inputs [42]. GFMs represent a further shift in which a reusable representation is learned before the downstream agricultural task [43]. Self-supervised learning is particularly compatible with Earth observation because satellite archives contain enormous quantities of imagery while expert-labelled agricultural variables remain comparatively scarce [44]. Nevertheless, the learned representation reflects the geographic, spectral, temporal, and preprocessing characteristics of the pretraining corpus, so apparent generality must be demonstrated empirically [45]. The key distinction is therefore not simply model size, but the separation of large-scale representation learning from task-specific supervision

How to cite this article: Wassaf A, Zeenat K. Geospatial Foundation Models for Intelligent Crop Monitoring: AI-Driven Multispectral Analysis, Yield Prediction, and Precision Agriculture. JOJ Horticulture & Arboriculture, 6(4). 555695. DOI: 10.19080/JOJHA.2026.06.555695

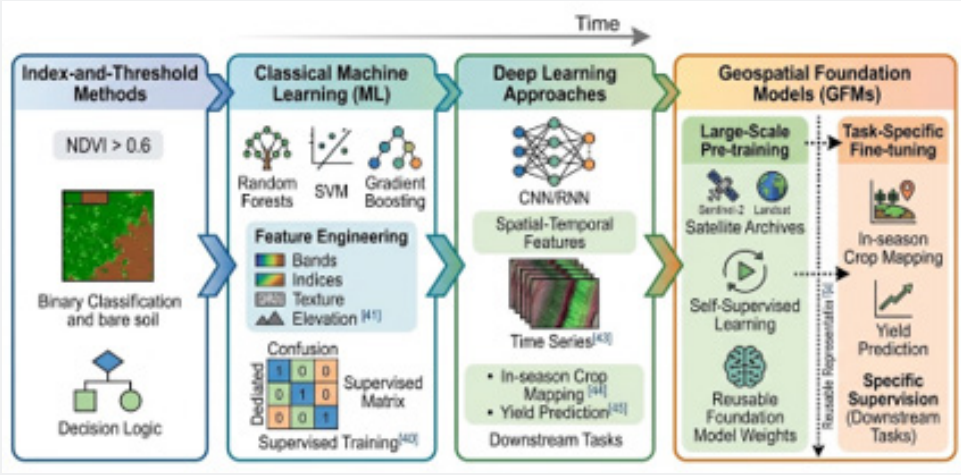


Figure 2: Evolution of analytical paradigms for multispectral crop monitoring, from index thresholds to geospatial foundation models.

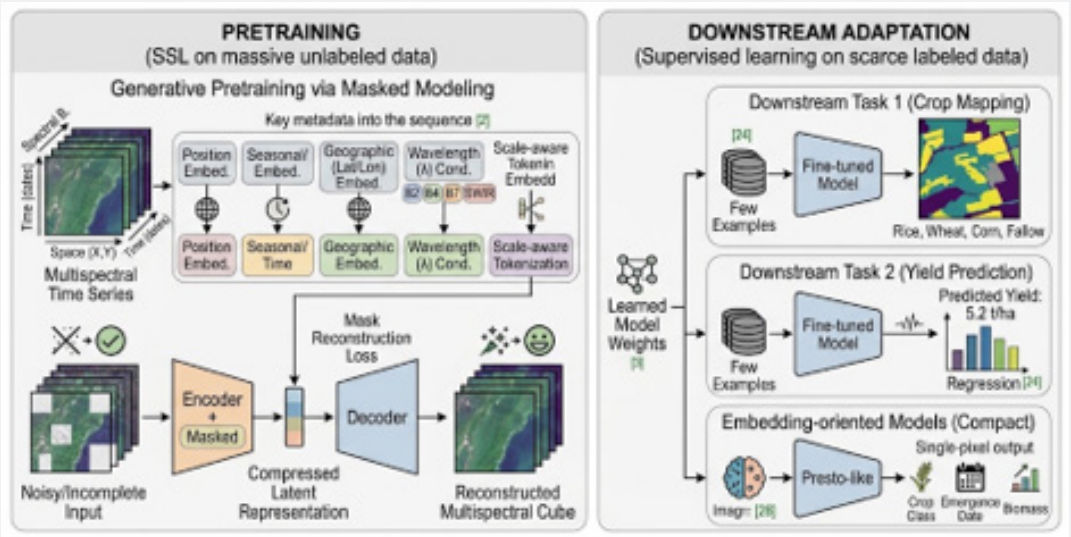


Figure 3: Generic masked-autoencoder pretraining for multispectral time series, and its downstream adaptation to crop tasks. Position, seasonal, geographic and wavelength embeddings give the model spectral, temporal and spatial context.

Data preprocessing and its consequences for pretrained models

Multispectral analysis is highly sensitive to preprocessing choices. Sentinel-2 observations may be supplied as top-of-atmosphere or surface-reflectance products, and atmospheric correction, cloud and shadow masking, temporal compositing, gap filling, geometric co-registration, and normalization can materially alter the inputs presented to a model [46]. HLS reduces part of this variability through harmonized processing across Landsat and Sentinel-2 and is therefore particularly useful for multi-temporal pretrained models [47]. This is especially important for agricultural studies that combine satellite archives with UAV observations or locally collected field data [48]. Differences in preprocessing can create

systematic distribution shifts that are easily mistaken for changes in crop condition or differences in model capability [49]. The HLS V2.0 product reduces cross-sensor inconsistency through atmospheric correction, cloud and shadow masking, bidirectional-reflectance normalization, bandpass adjustment, and common gridding [50]. UAV workflows introduce additional requirements, including reflectance-panel calibration, illumination correction, lens and geometric correction, and Ortho mosaicking.

How to cite this article: Wassaf A, Zeenat K. Geospatial Foundation Models for Intelligent Crop Monitoring: AI-Driven Multispectral Analysis, Yield Prediction, and Precision Agriculture. JOJ Horticulture & Arboriculture, 6(4). 555695. DOI: 10.19080/JOJHA.2026.06.555695

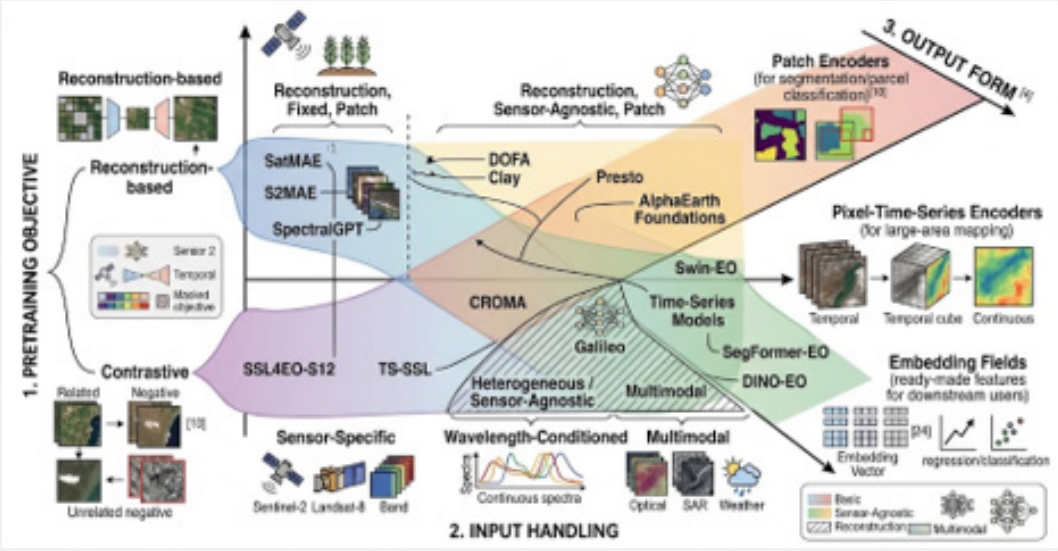


Figure 4: Taxonomy of geospatial foundation models for crop monitoring along three design axes. Placement of a model is indicative, since many models combine features from several branches.

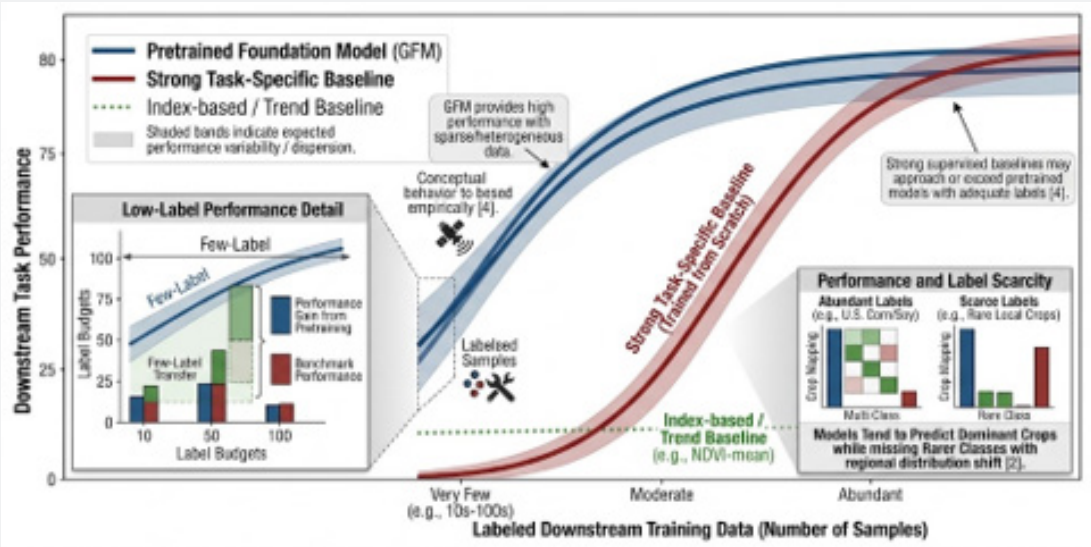


Figure 5: Conceptual label-efficiency hypothesis. This is an illustration of the expected behavior that empirical studies should test, and it does not depict measured results.

For foundation models, preprocessing is part of the model specification rather than a peripheral implementation detail. A pretrained encoder implicitly reflects the product level, spectral configuration, normalization statistics, masking policy, temporal sampling, and spatial scale of its pretraining data. [51] A downstream application that changes these assumptions without evaluation can confound model capability with input-distribution mismatch [25]. The analysis therefore recommend that GFM studies report the exact sensor and product level, band order, normalization procedure, cloud and shadow masking, temporal compositing, spatial resampling, and any transformations used before inference. [4] Even with harmonization, users should document scaling, missing-data treatment, temporal compositing, cloud screening, and normalization statistics because these operations define the effective input distribution seen by the encoder [4]. Reproducibility therefore requires reporting preprocessing as part of the model configuration rather than treating it as an undocumented data-engineering step. [4] This is especially important for agricultural studies that combine satellite archives with UAV observations or locally collected field data [1].

How to cite this article: Wassaf A, Zeenat K. Geospatial Foundation Models for Intelligent Crop Monitoring: AI-Driven Multispectral Analysis, Yield Prediction, and Precision Agriculture. JOJ Horticulture & Arboriculture, 6(4). 555695. DOI: 10.19080/JOJHA.2026.06.555695

Where possible, the downstream preprocessing pipeline should reproduce the conditions represented during pretraining.

Geospatial Foundation Models

What is different about Earth observation data

Foundation-model methods developed for natural images cannot be transferred to Earth observation without accounting for the distinctive structure of EO data. Satellite observations contain more channels than RGB imagery, spectral configurations differ among sensors, observations form irregular time series because of cloud and acquisition geometry, spatial resolution varies substantially, and reliable labels are often sparse and geographically clustered [2]. Agricultural scenes add another layer of complexity because crop identity and condition are strongly phenological: the same field may change substantially over a few weeks, while different fields may appear spectrally similar on a single date [2]. Self-supervised learning is particularly compatible with Earth observation because satellite archives contain enormous quantities of imagery while expert-labelled agricultural variables remain comparatively scarce [24]. Nevertheless, the learned representation reflects the geographic, spectral, temporal, and preprocessing characteristics of the pretraining corpus, so apparent generality must be demonstrated empirically [2]. An agricultural GFM therefore needs to represent spectral, spatial, temporal, and geographic context jointly.

Pretraining paradigms

Masked modelling adapts the masked-autoencoder paradigm to EO data by hiding a substantial fraction of input tokens and learning to reconstruct the missing content. Sat MAE introduced temporal embeddings and spectral grouping for multispectral and multi-temporal satellite imagery [2]. Prithvi extended masked modelling to multi-temporal HLS sequences, while Prithvi-EO-2.0 introduced larger models and explicit temporal and location embeddings [3]. Spectral and S2MAE explicitly model spectral–spatial structure, while Scale-MAE introduces scale-aware pretraining so that spatial scale becomes part of the learned representation [2]. Nevertheless, the learned representation reflects the geographic, spectral, temporal, and preprocessing characteristics of the pretraining corpus, so apparent generality must be demonstrated empirically [2]. Recent foundation-model research increasingly addresses these constraints through temporal embeddings, wavelength conditioning, multimodal alignment, and scale-aware encoders [28]. The agricultural question is consequently whether these architectural choices improve transfer under realistic label scarcity and distribution shift rather than whether these models simply increase benchmark accuracy [25]. Together, these approaches illustrate how the masking objective is being adapted to the physical structure of EO data rather than copied directly from natural-image pretraining.

Contrastive and distillation-based objectives learn representations by making related observations similar and unrelated observations distinguishable, rather than reconstructing pixels directly. SSL4EO-S12 provides multi-season Sentinel-1 and Sentinel-2 observations for large-scale self-supervised learning and has supported comparisons among contrastive, distillation, and masked objectives [1]. These approaches are useful when the desired representation should be invariant to nuisance variation such as illumination, seasonal acquisition differences, or sensor-specific appearance [4]. The principal advantage of this paradigm is not model size by itself but the possibility of learning reusable representations from observations for which dense agricultural labels are unavailable [4]. Self-supervised learning is particularly compatible with Earth observation because satellite archives contain enormous quantities of imagery while expert-labelled agricultural variables remain comparatively scarce [24]. Nevertheless, the learned representation reflects the geographic, spectral, temporal, and preprocessing characteristics of the pretraining corpus, so apparent generality must be demonstrated empirically [2]. Language-aligned learning is another emerging direction, but its value for agricultural decision support still requires domain-specific validation rather than assumptions based on natural-image systems.

Multimodal and sensor-flexible models attempt to learn representations that remain useful when optical, radar, hyperspectral, weather, terrain, or other observations are combined. CROMA jointly learns radar–optical representations using contrastive and masked reconstruction objectives [3]. DOFA uses wavelength-conditioned dynamic parameters to accommodate heterogeneous spectral configurations, while Clay incorporates sensor metadata and wavelength information into a masked-autoencoder framework [2]. Galileo extends multimodal representation learning across optical, radar, elevation, weather, and other inputs while explicitly modelling both global and local spatial structure [3]. Recent foundation-model research increasingly addresses these constraints through temporal embeddings, wavelength conditioning, multimodal alignment, and scale-aware encoders [28]. The agricultural question is consequently whether these architectural choices improve transfer under realistic label scarcity and distribution shift rather than whether these models simply increase benchmark accuracy [25]. The principal advantage of this paradigm is not model size by itself but the possibility of learning reusable representations from observations for which dense agricultural labels are unavailable [4]. These designs are particularly relevant to agriculture because cloud gaps, sensor changes, and environmental covariates can otherwise limit the continuity of optical-only monitoring.

Lightweight and embedding-oriented models address a different deployment constraint: the representation should

How to cite this article: Wassaf A, Zeenat K. Geospatial Foundation Models for Intelligent Crop Monitoring: AI-Driven Multispectral Analysis, Yield Prediction, and Precision Agriculture. JOJ Horticulture & Arboriculture, 6(4). 555695. DOI: 10.19080/JOJHA.2026.06.555695

be useful without requiring large-scale inference or task-specific training infrastructure. Presto operates on pixel-level time series and integrates optical, radar, weather, terrain, and location information in a compact transformer [10]. Alpha Earth Foundations produces annual, gridded embeddings that can be used as a reusable feature layer for downstream mapping [3]. Stalas Pretrain represents a complementary supervised route in which large labelled collections are used to learn general remote-sensing representations [4]. Self-supervised learning is particularly compatible with Earth observation because satellite archives contain enormous quantities of imagery while expert-labelled agricultural variables remain comparatively scarce [24]. Nevertheless, the learned representation reflects the geographic, spectral, temporal, and preprocessing characteristics of the pretraining corpus, so apparent generality must be demonstrated empirically [2]. Recent foundation-model research increasingly addresses these constraints through temporal embeddings, wavelength conditioning, multimodal alignment, and scale-aware encoders [28]. These approaches suggest that practical agricultural adoption may depend as much on inference cost, accessibility, and reproducibility as on peak benchmark accuracy.

Model landscape

(Table 5) highlights the transition from fixed-band and single-modality models toward sensor-flexible, multimodal, scale-aware, and embedding-oriented systems. Recent models increasingly encode wavelength, spatial scale, time, and geographic context explicitly, reflecting the fact that these variables are part of the measurement process rather than incidental metadata. At the same time, most GFMs were developed for general Earth-observation tasks rather than agriculture specifically. Consequently, crop mapping, stress monitoring, and yield estimation should be interpreted as downstream tests of general representations rather than evidence that a model has learned agronomic causality.

A taxonomy for agricultural use

For agricultural practitioners, three design axes are especially useful (Figure 4). The first is the pretraining objective: reconstruction-based models can preserve detailed spectral and spatial information, whereas contrastive approaches emphasis relationships that remain stable across augmented or temporally related observations [2]. The second is input handling: fixed-band models are straightforward but sensor-specific, while wavelength-conditioned and multimodal architectures are designed to accommodate heterogeneous inputs [10]. The third is output form: patch encoders are natural for segmentation, pixel-time-series encoders are attractive for large-area mapping, and embedding fields can provide a ready-made feature layer for users with limited machine-learning infrastructure [4]. Self-supervised learning is particularly compatible with Earth observation because satellite archives contain enormous quantities of imagery while expert-labelled agricultural variables remain comparatively scarce [24]. Nevertheless, the learned representation reflects the geographic, spectral, temporal, and preprocessing characteristics of the pretraining corpus, so apparent generality must be demonstrated empirically [2]. These axes describe trade-offs rather than a universal hierarchy; suitability depends on task requirements, label availability, spatial scale, temporal needs, and compute constraints.

Adapting a pretrained model to a crop task

Four adaptation strategies are commonly used. Frozen embeddings with a simple prediction head, such as linear regression, ridge regression, or gradient boosting, minimize the amount of labelled data and computation required and provide an important baseline [4]. Partial fine-tuning updates selected transformer blocks or a decoder, allowing more task-specific adaptation [4]. Parameter-efficient methods can reduce the number of trainable parameters when several regional or crop-specific variants are required [23]. Full fine-tuning provides the greatest flexibility but can overfit when agricultural labels are small or geographically concentrated [4]. Recent foundation-model research increasingly addresses these constraints through temporal embeddings, wavelength conditioning, multimodal alignment, and scale-aware encoders [28]. The agricultural question is consequently whether these architectural choices improve transfer under realistic label scarcity and distribution shift rather than whether these models simply increase benchmark accuracy [25]. A rigorous study should compare at least one frozen or parameter-efficient setting with a from-scratch baseline so that the contribution of pretraining is measurable rather than inferred.

Task heads should reflect the statistical structure of the agricultural target. Segmentation heads are appropriate for parcel and pixel maps, temporal classifiers are suited to crop-type or phenology sequences, and regression heads can be applied to field-level yield or biophysical variables [5]. For yield estimation, imagery should normally be combined with covariates that are only weakly observable from reflectance, including weather, soil, cultivar, management, irrigation, and planting date [5]. Recent systematic reviews show that remote-sensing yield estimation continues to be constrained by limited ground truth, inconsistent data quality, temporal gaps, and difficulty transferring models between regions [5]. Consequently, an embedding extracted from reflectance should normally be combined with weather, soil, management, and crop-calendar variables rather than interpreted as a complete representation of yield formation [5]. The strongest evaluation design should test both predictive accuracy and the stability of the prediction under geographic, seasonal, and management changes [25]. Treating imagery alone as a complete representation of yield formation can therefore impose an unrealistic causal assumption.

How to cite this article: Wassaf A, Zeenat K. Geospatial Foundation Models for Intelligent Crop Monitoring: AI-Driven Multispectral Analysis, Yield Prediction, and Precision Agriculture. JOJ Horticulture & Arboriculture, 6(4). 555695. DOI: 10.19080/JOJHA.2026.06.555695

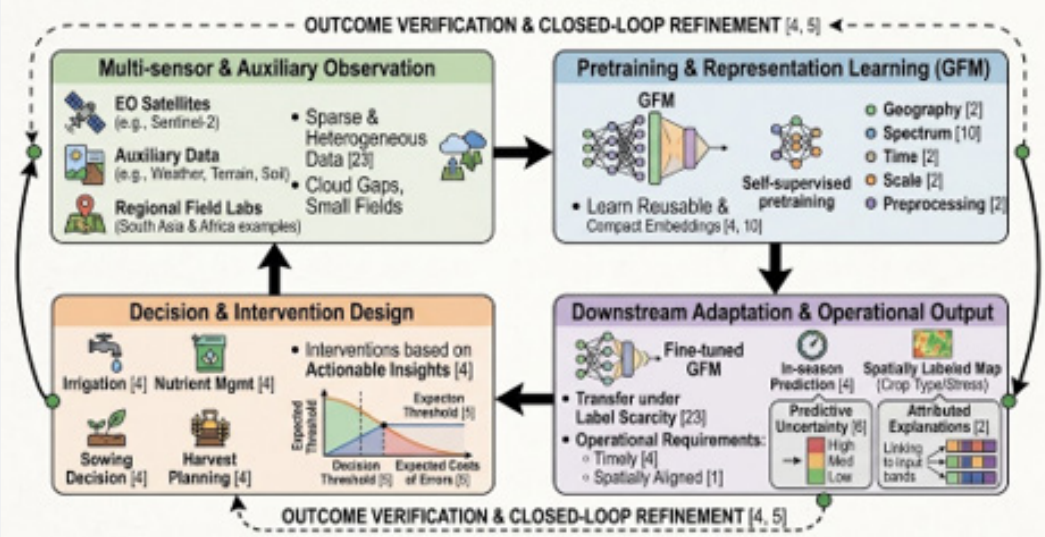


Figure 6: Insight-to-action framework for GFM-enabled crop management. The dashed loop shows that verified outcomes should refine labels, models and recommendations.

Table 3: Widely used multispectral indices, expressed in Sentinel-2 band notation.

Index	Formula	Sensitivity	Typical use
NDVI [29]	(B8 - B4) / (B8 + B4)	Green biomass, cover	Crop vigour, phenology, yield proxy
EVI [29]	2.5 (B8 - B4) / (B8 + 6 B4 - 7.5 B2 + 1)	Biomass; reduced saturation	Dense canopies, atmosphere correction
SAVI [29]	1.5 (B8 - B4) / (B8 + B4 + 0.5)	Cover with soil background	Early season, sparse canopies
GNDVI [29]	(B8 - B3) / (B8 + B3)	Chlorophyll	Nitrogen status
NDRE [29]	(B8 - B5) / (B8 + B5)	Chlorophyll (red edge)	Mid-to-late-season nitrogen, stress
NDWI/NDMI [29]	(B8A - B11) / (B8A + B11)	Canopy water content	Water stress, irrigation

Table 4: Qualitative comparison of analytical paradigms.

Dimension	Indices / thresholds	Classical ML	Deep learning	Foundation models
Feature design	Manual, fixed	Manual, flexible	Learned per task	Learned once, reused
Label demand	Very low	Moderate	High	Low after pretraining (task-dependent)
Temporal modelling	Simple statistics	Aggregated features	RNN / attention	Temporal tokens and embeddings
Cross-region transfer	Poor	Limited	Fragile	Promising but under-tested
Compute cost	Negligible	Low	Moderate-high	Very high to pretrain; low to reuse
Interpretability	High	Moderate	Low	Low; needs dedicated tools

Applications in Crop Monitoring and Yield Estimation

Crop-type and cropland mapping

Crop-type and cropland mapping is currently the agricultural application most directly represented in GFM evaluations. Time-series Sentinel-2 studies have demonstrated that crop identity can be inferred from temporal patterns rather than single-date spectra, including in-season mapping settings where labels are limited. [1] PASTIS provides parcel-level Sentinel-2 time-series annotations for reproducible temporal segmentation, while Crop Harvest provides geographically diverse crop labels for low-label evaluation [9]. GFMs have subsequently been evaluated in both dense-label and few-label settings: Prithvi has been fine-tuned for multi-temporal crop segmentation, while Presto has been assessed on Crop Harvest-style tasks [3]. Evaluation design is therefore part of the scientific contribution because random splits can place spatially adjacent or temporally related observations in both training and test sets [25]. Benchmarks such as GEO-Bench, PhilEO, and PANGAEA were developed partly to improve consistency in downstream evaluation and to expose differences among models that may be hidden by single-task experiments. These results support the feasibility of transfer learning, but these models do not remove the need for regional validation because crop calendars, class definitions, field geometry, and label quality

How to cite this article: Wassaf A, Zeenat K. Geospatial Foundation Models for Intelligent Crop Monitoring: AI-Driven Multispectral Analysis, Yield Prediction, and Precision Agriculture. JOJ Horticulture & Arboriculture, 6(4). 555695. DOI: 10.19080/JOJHA.2026.06.555695

differ substantially across agricultural systems [25]. A 2026 global agriculture-mapping foundation model, Agri FM, was pretrained on more than 25 million samples from MODIS, Landsat-8/9, and Sentinel-2 and was evaluated for agricultural land mapping, field-boundary delineation, agricultural land-use mapping, and specific crop mapping [9]. This result expands the evidence base from single-task crop classification toward unified multi-source and multi-temporal agricultural mapping.

Table 5: Representative geospatial foundation models relevant to multispectral crop monitoring. Descriptions summaries the cited publications and project documentation.

Model	Year	Objective / architecture	Pretraining data	Inputs
Sat MAE [16]	2022	Masked autoencoder Vitt with spectral-group and temporal encodings	Multispectral fMoW-derived data	Multispectral, multi-temporal
SSL4EO-S12 baselines	2022-23	MoCo, DINO, MAE, data2vec pretraining	~250k locations, four seasons, Sentinel-1/2	Sentinel-2, Sentinel-1
Prithvi-EO-1.0 [18]	2023	Temporal Vitt-MAE with 3D patch embedding	HLS over the contiguous United States	6 HLS bands, up to 3 dates
CROMA [17]	2023	Contrastive plus masked reconstruction, radar-optical	SSL4EO-S12	Sentinel-1 and Sentinel-2
Presto [34]	2023	Lightweight transformer on pixel time series	Global multi-source pixel time series	Optical, radar, weather, terrain, location
SatlasPretrain [23]	2023	Supervised multi-task pretraining	Large multi-task labelled collection	Sentinel-2 and high-resolution imagery
Spectral [22]	2023	3D masked autoencoder, spectral-spatial tokens	Large Sentinel-2 collection	Multispectral (variable bands)
DOFA [19]	2024	Wavelength-conditioned dynamic hypernetwork	Multiple sensors (optical, radar, hyperspectral)	Any band set
Prithvi-EO-2.0 [18]	2024	Temporal Vitt-MAE with time and location embeddings; 300M and 600M variants	Global HLS	6 HLS bands, multi-temporal
Clay v1 [20]	2024	MAE with wavelength-aware embeddings	Sentinel-2, Sentinel-1, Landsat, NAIP and others	Any supported band set
Galileo [21]	2025	Multimodal, multi-scale global-local objectives	Global multi-source data	Optical, radar, weather, terrain; images and series
Alpha Earth Foundations [6]	2025	Embedding-field model, annual 64-dimensional embeddings	Multi-source EO, climate and other data	Precomputed 10 m embeddings
Agri FM [52]	2026	Multi-source, multi-temporal foundation model with synchronized spatio-temporal down sampling	Global dataset with >25 million MODIS, Landsat-8/9 and Sentinel-2 samples	Optical multispectral, multi-source, multi-temporal

Table 6: Benchmarks and datasets relevant to agricultural evaluation of geospatial foundation models.

Resource	Type	Content	Relevance to agriculture
GEO-Bench [25]	Multi-task benchmark	Suite of EO classification and segmentation tasks	General transfer assessment; few crop-specific tasks
PANGAEA [27]	GFM benchmark	Diverse datasets, several sensors and tasks, including crop segmentation	Compares GFMs with supervised baselines
Crop Harvest [9]	Labelled points	Global crop-type and cropland labels with paired time series	Few-label, cross-region mapping
PASTIS [29]	Parcel segmentation	Sentinel-2 time series with parcel-level crop labels (France)	Temporal crop segmentation
Crop Net [5]	Yield dataset	Sentinel-2, weather and county-level yields (US)	Yield prediction with climate covariates
SSL4EO-S12 [3]	Pretraining corpus	Multi-season Sentinel-1/2 imagery at ~250k locations	Pretraining and ablation of objectives
Bequeathment [36]	Scene classification	Sentinel-2 patches with multi-label land cover	Indirect: agricultural land-cover classes
Eurostat [36]	Scene classification	Sentinel-2 patches, ten land-use classes	Indirect: annual and permanent crop classes
2026 agricultural GFM benchmark [48]	Agriculture-specific GFM benchmark	Regional transfer tests for Prithvi, Spectral and Sat MAE across four U.S. states	Direct assessment of geographic transfer and rare-crop robustness

How to cite this article: Wassaf A, Zeenat K. Geospatial Foundation Models for Intelligent Crop Monitoring: AI-Driven Multispectral Analysis, Yield Prediction, and Precision Agriculture. JOJ Horticulture & Arboriculture, 6(4). 555695. DOI: 10.19080/JOJHA.2026.06.555695

Table 7: Main challenges for GFMs in crop monitoring and emerging mitigations.

Challenge	Why it matters for agriculture	Emerging mitigation
Label scarcity and noise	Ground truth on yield, stress and crop type is costly; administrative labels are noisy	Self-supervised pretraining; weak supervision; active labelling; crowdsourced field data
Domain shift	Cropping calendars, varieties and management differ by region and year	Location and time embeddings; leave-region-out evaluation; regional fine-tuning
Small fields, coarse pixels	Smallholder fields are often smaller than a 10 m pixel or a model patch	Higher-resolution inputs; scale-aware models; UAV-satellite fusion
Clouds and irregular series	Growing seasons in monsoon regions are cloudy, leaving gaps at critical stages	Radar-optical models; gap-tolerant temporal encoders
Sensor heterogeneity	Farms and projects use different satellites and UAV cameras	Wavelength-conditioned encoders; harmonized products such as HLS
Evaluation leakage	Spatial and temporal autocorrelation inflates reported accuracy	Blocked cross-validation; hold-out years and regions; strong baselines
Interpretability	Agronomists and farmers need actionable, credible explanations	Attribution methods; hybrid models with physical or crop-model constraints
Compute and access	Large models are costly to pretrain and to run	Lightweight models; distillation; precomputed embeddings; open checkpoints
Lack of yield-specific pretraining	Generic objectives do not encode phenology-yield relationships	Phenology-aware objectives; pretraining with crop-model simulations

Phenology stress and biophysical variables

Phenology, water status, nutrient status, and other biophysical variables are closer to management decisions than generic land-cover classes, but these models are also harder to label because reliable reference data often require field campaigns or calibrated proximal measurements. Multi-temporal models with explicit time and location representations are well suited to phenological structure [4]. Red-edge and SWIR observations provide complementary information on chlorophyll, canopy structure, and water status, while recent optical remote-sensing reviews continue to identify red-edge information as important for crop stress and biochemical retrieval [2].

Nevertheless, the published GFM literature remains dominated by classification and segmentation benchmarks; direct evaluations against field-measured stress, LAI, chlorophyll, or nutrient status are comparatively limited. Physics-based models such as PROSAIL could provide synthetic training examples or physically constrained objectives, but this integration remains an open research direction [2]. A 2026 multimodal study reported that its agriculture-focused foundation model detected crop stress and yield-failure risk earlier than a satellite-only baseline, with the largest retrospective gains occurring during compound heat and drought events [5]. These findings strengthen the case for evaluating temporal lead time and uncertainty rather than relying only on end-of-season classification accuracy.

Yield estimation

Satellite-based yield estimation has a substantial methodological literature, but recent reviews continue to identify limitations in label availability, temporal gaps, spatial transfer, and the integration of environmental drivers. Modern datasets increasingly combine satellite imagery with weather and administrative yield information; Crop Net, for example, aligns Sentinel-2 imagery, meteorological variables, and USDA yield data across more than 2,200 U.S. counties over multiple years [5]. Recent evidence also shows that uncertainty estimation can be incorporated explicitly, rather than reporting only a point prediction [2]. Consequently, an embedding extracted from reflectance should normally be combined with weather, soil, management, and crop-calendar variables rather than interpreted as a complete representation of yield formation [5]. The strongest evaluation design should test both predictive accuracy and the stability of the prediction under geographic, seasonal, and management changes [25]. A yield model becomes operationally meaningful only when its forecast is sufficiently early, spatially resolved, and calibrated to support a specific agricultural decision. These developments provide a useful foundation for testing whether GFM embeddings add value beyond vegetation indices, conventional machine learning, and process-based crop modelling [5]. A 2026 systematic review of Sentinel-2 yield estimation identified empirical machine-learning approaches, crop-model data assimilation, and Sentinel-1/Sentinel-2 fusion as the principal methodological pathways, while highlighting persistent constraints from ground-truth scarcity, cloud gaps, and cross-region transfer [5]. The 2026 evidence therefore reinforces the need to evaluate foundation-model representations against established yield-estimation pipelines rather than treating pretraining as a substitute for agronomic observations.

GFMs can enter yield-estimation pipelines in several ways. First, frozen embeddings can be extracted for fields or pixels

How to cite this article: Wassaf A, Zeenat K. Geospatial Foundation Models for Intelligent Crop Monitoring: AI-Driven Multispectral Analysis, Yield Prediction, and Precision Agriculture. JOJ Horticulture & Arboriculture, 6(4). 555695. DOI: 10.19080/JOJHA.2026.06.555695

and combined with weather, soil, management, and other agronomic covariates in a lightweight regression model [10]. Second, a pretrained encoder can be fine-tuned using field-level yield observations when sufficient labels are available [5]. Third, embedding products such as Alpha Earth can be combined with crop-growth or other agronomic models to provide spatially consistent learned covariates [3]. A central methodological requirement is a fair comparison with strong non-GFM baselines. [4] To date, the literature does not provide a sufficiently broad, multi-country evaluation showing that modern GFM embeddings consistently outperform carefully constructed index-based, supervised, or process-informed yield models under spatially and temporally independent testing [5]. Yield is a particularly demanding target because it integrates weather, soil, cultivar, management, pests, water availability, and accumulated crop development over an entire season [5]. Recent systematic reviews show that remote-sensing yield estimation continues to be constrained by limited ground truth, inconsistent data quality, temporal gaps, and difficulty transferring models between regions [5]. This remains a major evidence gap.

Evaluation design is particularly important for yield because observations are spatially and temporally autocorrelated. Randomly assigning neighboring fields or adjacent years to training and test sets can allow information leakage and produce overly optimistic estimates [4]. Stronger protocols use spatial blocking, leave-one-year-out or leave-region-out testing, and explicit reporting of the geographic and temporal separation between training and evaluation data [25]. Baselines should include simple predictors such as regional means or trends, conventional vegetation-index models, and strong task-specific machine-learning models [4]. Label-efficiency curves are more informative than a single low-label operating point because this information [11]. This evidence makes scale bridging a practical evaluation problem rather than only an architectural objective.

UAV and satellite scale bridging

UAV multispectral imagery captures sub-field and plot-scale heterogeneity that is usually unresolved by 10–30 m satellite products, while satellites provide the spatial coverage required for regional monitoring. This scale mismatch is a major challenge for agricultural GFM deployment [24]. Scale-aware pretraining and multimodal architectures address parts of the problem, but direct evidence for transferring representations between UAV and satellite imagery remains limited [2]. Additional challenges include differences in spectral response functions, radiometric calibration, viewing geometry, illumination, Ortho mosaic quality, and mixed pixels at satellite resolution [2]. This requirement motivates recent work on wavelength-aware and scale-aware representations that attempt to make transfer across heterogeneous observations more systematic [24]. The practical consequence is that observations from different platforms cannot be treated as interchangeable without accounting for spatial support, spectral response functions, radiometric calibration, and temporal sampling [2]. Harmonized products reduce some of these differences, but these models do not eliminate the information loss that occurs when fine-scale field structure is aggregated to a coarser ground sampling distance.

A robust UAV–satellite framework therefore needs explicit treatment of both spatial scale and sensor physics rather than simple image resizing. [1] Recent 2026 agricultural foundation-model studies also show that high-resolution foundation-model adaptation can support fragmented field delineation, including smallholder fields below 0.1 ha, although performance remains dependent on inference strategy, image resolution, and temporal models are summarized from the cited publications [4]. Numerical performance values are intentionally not reproduced because the studies use different datasets, splits, metrics, and baselines; instead, the cases are used to identify transferable methodological lessons and limitations [4]. The corresponding differences in spatial representation, temporal resolution, pretraining data, and computational requirements are directly relevant to agricultural use because crop-monitoring tasks operate at different spatial and temporal scales [3]. The reported findings should therefore be interpreted within the original datasets and evaluation protocols rather than generalized across all crops and regions [25]. Independent agricultural replication is especially important because many published evaluations remain concentrated in regions with comparatively strong satellite coverage and labelled datasets [25]. (Table 6) provides a concise comparison.

Case Studies

The three cases below represent distinct points in the current GFM design space: a patch-based temporal encoder adapted to crop segmentation, a lightweight pixel-time-series model for few-label global mapping, and an embedding-field approach for sparse-label mapping. These representation to downstream EO tasks, including crop segmentation, and subsequent work expanded the Prithvi family to larger models with explicit temporal and geographic embeddings [3]. These case studies illustrate complementary deployment strategies rather than a universal hierarchy among foundation models [3]. The corresponding differences in spatial representation, temporal resolution, pretraining data, and computational requirements are directly relevant to agricultural use because crop-monitoring tasks operate at different spatial and temporal scales [3]. The reported findings should therefore be interpreted within the original datasets and evaluation protocols rather than generalized across all crops and regions [25]. Independent agricultural replication is especially important because many published evaluations remain concentrated in regions with comparatively strong satellite coverage and labelled datasets [25]. The main contribution of this case is methodological: it demonstrates a complete pretraining-to-agriculture pipeline using multispectral time series rather than treating crop mapping as a static image-classification problem.

How to cite this article: Wassaf A, Zeenat K. Geospatial Foundation Models for Intelligent Crop Monitoring: AI-Driven Multispectral Analysis, Yield Prediction, and Precision Agriculture. JOJ Horticulture & Arboriculture, 6(4). 555695. DOI: 10.19080/JOJHA.2026.06.555695

Prithvi-EO for multi-temporal crop segmentation

Context and approach: Prithvi-EO-1.0 was pretrained with masked reconstruction on HLS imagery and subsequently adapted to several EO tasks, including multi-temporal crop segmentation. The temporal formulation is important because crop classes are often distinguished by phenological trajectories rather than by a single-date spectral signature. The reported findings should therefore be interpreted within the original datasets and evaluation protocols rather than generalized across all crops and regions. Independent agricultural replication is especially important because many published evaluations remain concentrated in regions with comparatively strong satellite coverage and labelled datasets. Taken together, the cases support continued investigation while also showing why model transfer must be evaluated explicitly. These case studies illustrate complementary deployment strategies rather than a universal hierarchy among foundation models. The case demonstrates how a general pretrained encoder can be adapted to a crop-specific segmentation problem with substantially less task-specific representation learning than a model trained entirely from scratch.

Findings: The published evaluation reports useful transfer from the pretrained also require explicit treatment during adaptation and evaluation representation to downstream EO tasks, including crop segmentation, and subsequent work expanded the Prithvi family to larger models with explicit temporal and geographic embeddings. These case studies illustrate complementary deployment strategies rather than a universal hierarchy among foundation models. The corresponding differences in spatial representation, temporal resolution, pretraining data, and computational requirements are directly relevant to agricultural use because crop-monitoring tasks operate at different spatial and temporal scales. The reported findings should therefore be interpreted within the original datasets and evaluation protocols rather than generalized across all crops and regions. Independent agricultural replication is especially important because many published evaluations remain concentrated in regions with comparatively strong satellite coverage and labelled datasets. The main contribution of this case is methodological: it demonstrates a complete pretraining-to-agriculture pipeline using multispectral time series rather than treating crop mapping as a static image-classification problem.

Lessons and caveats: The evaluation setting remains favorable because the pretraining data, downstream imagery, and labels are comparatively consistent in sensor and geographic context [3]. Consequently, the case establishes feasibility but does not by itself establish transfer to regions with different crop calendars, field sizes, sensor characteristics, or cloud regimes [4]. Independent agricultural replication is especially important because many published evaluations remain concentrated in regions with comparatively strong satellite coverage and labelled datasets [25]. Taken together, the cases support continued investigation while also showing why model transfer must be evaluated explicitly [4]. These case studies illustrate complementary deployment strategies rather than a universal hierarchy among foundation models [3]. The corresponding differences in spatial representation, temporal resolution, pretraining data, and computational requirements are directly relevant to agricultural use because crop-monitoring tasks operate at different spatial and temporal scales [3]. Class imbalance and the unequal representation of minor crops deployment strategies rather than a universal hierarchy among foundation models [3]. It therefore highlights model efficiency and data design as practical dimensions of GFM usefulness, alongside predictive accuracy.

Lessons and caveats: Pixel-level processing does not explicitly model neighborhood structure, field geometry, or spatial texture. This can be limiting in fragmented landscapes where neighboring pixels provide information needed to distinguish crops. These case studies illustrate complementary deployment strategies rather than a universal hierarchy among foundation models. The corresponding differences in spatial representation, temporal resolution, pretraining data, and computational requirements are directly relevant to agricultural use because crop-monitoring tasks operate at different spatial and temporal scales. The reported findings should therefore be interpreted within the original datasets and evaluation protocols rather than generalized across all crops and regions. [Independent agricultural replication is especially important because many published evaluations remain concentrated in regions with comparatively strong satellite coverage and labelled datasets. Reported performance also depends on the selected tasks and validation splits; broader testing across crops, seasons, field sizes, and geographic regions is necessary before generalizing the findings to agricultural deployment.

Presto and few-label crop mapping on Crop Harvest

Context and approach: Crop Harvest contains more than 90,000 geographically diverse agricultural samples and was designed to expose spatial, semantic, and temporal domain-shift challenges in crop classification [4]. Presto is a compact transformer pretrained on pixel-level remote-sensing time series that combine optical, radar, weather, terrain, and location information [10]. Taken together, the cases support continued investigation while also showing why model transfer must be evaluated explicitly [4]. These case studies illustrate complementary deployment strategies rather than a universal hierarchy among foundation models [3]. The corresponding differences in spatial representation, temporal resolution, pretraining data, and computational requirements are directly relevant to agricultural use because crop-monitoring tasks operate at different spatial and temporal scales [3]. The reported findings should therefore be interpreted within the original datasets and evaluation protocols rather than generalized across all crops and regions [25]. Because it operates on time-

How to cite this article: Wassaf A, Zeenat K. Geospatial Foundation Models for Intelligent Crop Monitoring: AI-Driven Multispectral Analysis, Yield Prediction, and Precision Agriculture. JOJ Horticulture & Arboriculture, 6(4). 555695. DOI: 10.19080/JOJHA.2026.06.555695

series features rather than requiring image patches or explicit field boundaries, it offers a computationally efficient alternative for large-area mapping.

Findings. The Presto study reports competitive performance across globally distributed remote-sensing tasks while using substantially fewer computational resources than many larger models [4]. The case is important for agriculture because label scarcity and limited computational capacity can coexist, particularly in regional and smallholder applications [11]. The reported findings should therefore be interpreted within the original datasets and evaluation protocols rather than generalized across all crops and regions [25]. Independent agricultural replication is especially important because many published evaluations remain concentrated in regions with comparatively strong satellite coverage and labelled datasets [25]. Taken together, the cases support continued investigation while also showing why model transfer must be evaluated explicitly [4]. These case studies illustrate complementary strong satellite coverage and labelled datasets [25]. Taken together, the cases support continued investigation while also showing why model transfer must be evaluated explicitly [4]. For agriculture, the practical proposition is that users can obtain a reusable feature layer and train a lightweight classifier or regressor with relatively few local labels, without constructing a complete satellite-processing pipeline.

Findings. The developers report strong performance across a diverse set of mapping evaluations, including tasks involving land use and crop type, particularly under sparse-label conditions [9]. These results indicate the potential value of embedding fields as an infrastructure layer for downstream agricultural mapping [4]. Taken together, the cases support continued investigation while also showing why model transfer must be evaluated explicitly [4]. These case studies illustrate complementary deployment strategies rather than a universal hierarchy among foundation models [3]. The corresponding differences in spatial representation, temporal resolution, pretraining data, and computational requirements are directly relevant to agricultural use because crop-monitoring tasks operate at different spatial and temporal scales [3]. The reported findings should therefore be interpreted within the original datasets and evaluation protocols rather than generalized across all crops and regions. However, the evaluation is primarily a general mapping assessment, so agricultural replication by independent groups remains important before extrapolating the results to field-level yield or management decisions [5]. A 2026 independent benchmark of Alpha Earth embeddings found strong performance on several agricultural tasks when models were trained with local data, while also identifying limited spatial transferability, interpretability, and temporal sensitivity as important constraints [3]. This independent evidence materially strengthens the review's earlier caution that embedding-field models should not be assumed to generalize uniformly across agricultural regions.

Lessons and caveats: Annual embeddings compress information over time and may therefore lose short-lived signals associated with within-season stress, irrigation events, or rapid phenological transitions [2]. These models may be useful as covariates for end-of-season

Embedding fields for sparse-label mapping

Context and approach: Alpha Earth Foundations generates annual geospatial embeddings by integrating spatial, temporal, and measurement context from multiple Earth-observation sources [3]. The corresponding differences in spatial representation, temporal resolution, pretraining data, and computational requirements are directly relevant to agricultural use because crop-monitoring tasks operate at different spatial and temporal scales [3]. The reported findings should therefore be interpreted within the original datasets and evaluation protocols rather than generalized across all crops and regions [25]. Independent agricultural replication is especially important because many published evaluations remain concentrated in regions with comparatively mapping and yield models, whereas near-real-time monitoring requires temporally resolved representations [5]. The reported findings should therefore be interpreted within the original datasets and evaluation protocols rather than generalized across all crops and regions [25]. Independent agricultural replication is especially important because many published evaluations remain concentrated in regions with comparatively strong satellite coverage and labelled datasets [25]. Taken together, the cases support continued investigation while also showing why model transfer must be evaluated explicitly [4]. These case studies illustrate complementary deployment strategies rather than a universal hierarchy among foundation models [3]. As with the other cases, the evidence for direct yield enhancement remains indirect because an accurate embedding or map is not equivalent to a verified agronomic benefit.

Benchmarks and Evaluation Practice

Reliable comparison requires shared benchmarks with consistent task definitions, splits, metrics, and reporting standards. GEO-Bench and PANGAEA broaden evaluation beyond individual model papers, while PhilEO provides an additional framework for controlled comparison of EO foundation models. [3] PANGAEA's subsequent evaluation evidence indicates that GFMs do not consistently outperform strong supervised baselines across all tasks, reinforcing the need for model-specific rather than universal claims (Figure 5) [2]. Agriculture-specific resources remain fragmented: Crop Harvest targets global few-label crop mapping, PASTIS supports parcel-level temporal segmentation, and Crop Net targets multimodal yield prediction with weather information [5]. General scene datasets such as Euro SAT and Big Earth Net can be useful for representation learning, but these models only partially reflect agricultural decision problems [5] [4]. These principles are necessary for distinguishing genuine transfer

How to cite this article: Wassaf A, Zeenat K. Geospatial Foundation Models for Intelligent Crop Monitoring: AI-Driven Multispectral Analysis, Yield Prediction, and Precision Agriculture. JOJ Horticulture & Arboriculture, 6(4). 555695. DOI: 10.19080/JOJHA.2026.06.555695

from performance that arises from favorable data overlap or task-specific adaptation [10]. Evaluation design is therefore part of the scientific contribution because random splits can place spatially adjacent or temporally related observations in both training and test sets. A widely adopted, multi-country benchmark focused on field-level yield, stress, and management-relevant outcomes is still lacking [5]. A 2026 controlled benchmark of Prithvi, Spectral GPT, and Sat MAE found substantial degradation under regional distribution shift across four U.S. states, with models tending to predict dominant crops while missing rarer classes (Table 6) [2]. The result provides direct evidence that region-aware evaluation should be treated as a core benchmark requirement for agricultural GFMs.

A central hypothesis is that pretrained representations should become useful with fewer task-specific labels than models trained from scratch, particularly when the downstream data are sparse or heterogeneous. Figure 6 illustrates this hypothesis conceptually rather than empirically [4]. In practice, the label-efficiency curve can vary substantially by crop, region, sensor, task, and model, and strong supervised baselines may approach or exceed pretrained models when adequate labels are available [4]. Label-efficiency curves are more informative than a single low-label operating point because these models reveal how rapidly a foundation representation becomes useful as supervision increases [4]. These principles are necessary for distinguishing genuine transfer from performance that arises from favorable data overlap or task-specific adaptation [10]. Evaluation design is therefore part of the scientific contribution because random splits can place spatially adjacent or temporally related observations in both training and test sets [25]. Studies should therefore report complete label-efficiency curves, repeated trials, uncertainty or dispersion, and the exact data-splitting protocol rather than a single favorable operating point.

Recommended practice includes that agricultural GFM studies adopt at least the following practices: spatially blocked validation and leave-one-year-out or leave-one-region-out testing; comparison with regional-mean, trend, vegetation-index, and strong task-specific baselines; calibration or prediction-interval assessment; explicit disclosure of pretraining-data overlap with test regions; evaluation across multiple label budgets; and reporting of training, adaptation, and inference costs. Agricultural evaluation requires an additional emphasis on crop calendars, field boundaries, yield measurement procedures, and the geographic distribution of labels [5]. Label-efficiency curves are more informative than a single low-label operating point because these models reveal how rapidly a foundation representation becomes useful as supervision increases [4]. These principles are necessary for distinguishing genuine transfer from performance that arises from favorable data overlap or task-specific adaptation [10]. Evaluation design is therefore part of the scientific contribution because random splits can place spatially adjacent or temporally related observations in both training and test sets [25]. These practices make it easier to distinguish genuine transfer from data

From Prediction to Yield Enhancement

Most GFM studies evaluate predictive performance, whereas yield enhancement is ultimately a decision problem. Useful outputs must support actions such as irrigation scheduling, nutrient management, crop scouting, sowing decisions, or harvest planning. A prediction has practical value only when it is available early enough to influence the decision, is spatially aligned with the management unit, carries an interpretable uncertainty estimate, and is sufficiently credible for the intended user [6]. Consequently, validation should include decision thresholds, expected costs of errors, and measurable changes in yield, input use, or economic return where interventions are feasible [5]. The transition from representation quality to agronomic value should therefore be treated as an empirical validation problem [4]. This distinction matters because predictive performance and intervention effectiveness answer different scientific questions [4]. Figure 6 therefore frames GFM deployment as a closed loop from observation and prediction to decision, outcome measurement, and model refinement. Four requirements follow. First, lead time matters: in-season forecasts can inform management, whereas end-of-season estimates have different uses such as reporting, insurance, or marketing. [4]. Second, spatial resolution must match the management unit; variable-rate interventions may require UAV or very-high-resolution satellite observations in addition to medium-resolution imagery.

[1]. Third, uncertainty and explanation should be treated as model outputs rather than afterthoughts; attribution methods can help relate predictions to spectral or temporal features, but transformer embeddings remain harder to interpret agronomically than simple indices [2]. Fourth, outcome verification is essential [4]. Recent crop-yield reviews similarly emphasize the importance of timely forecasts, reliable observations, and integration of environmental covariates when models are intended for operational use. [5]. Consequently, validation should include decision thresholds, expected costs of errors, and measurable changes in yield, input use, or economic return where interventions are feasible. Evidence that a model improves yield, reduces input use, or changes farm-level economic outcomes must come from field experiments, matched comparisons, or other credible intervention designs rather than from prediction accuracy alone [5]. The 2026 Sentinel-2 yield-estimation literature further shows that operational performance depends on combining high-resolution imagery with temporal information, environmental covariates, and appropriate validation rather than relying on a single spectral representation [5]. Accordingly, yield enhancement should be evaluated as a decision-support problem with measurable agronomic outcomes. Smallholder regions in South Asia and sub-Saharan Africa illustrate the combined challenge of label scarcity, small fields, cloud-related observation gaps, heterogeneous management, and limited computational infrastructure. These conditions do not automatically make GFMs

How to cite this article: Wassaf A, Zeenat K. Geospatial Foundation Models for Intelligent Crop Monitoring: AI-Driven Multispectral Analysis, Yield Prediction, and Precision Agriculture. JOJ Horticulture & Arboriculture, 6(4). 555695. DOI: 10.19080/JOJHA.2026.06.555695

more effective; rather, these models create a demanding test of whether label-efficient and sensor-flexible representations can transfer beyond the regions that dominate existing EO datasets [23]. This distinction matters because predictive performance and intervention effectiveness answer different scientific questions [4]. A model can achieve low prediction error while providing insufficient lead time, inadequate spatial resolution, or poorly calibrated uncertainty for a real management decision [6]. Recent crop-yield reviews similarly emphasize the importance of timely forecasts, reliable observations, and integration of environmental covariates when models are intended for operational use [5]. Multimodal optical–radar models, lightweight architectures, precomputed embeddings, and locally collected field observations are therefore important candidates for evaluation in these settings.

Challenges and Limitations

Consolidates (Table 7) the principal technical and operational challenges. Some are common to EO foundation models, including domain shift, benchmark inconsistency, interpretability, and computational cost; others are especially important in agriculture, including small fields, noisy yield labels, irregular crop calendars, and the absence of large-scale agronomic supervision. Computational requirements can further restrict reproducibility when the original pretraining environment cannot be reproduced by agricultural institutions or researchers with limited infrastructure. Interpretability is another important consideration because high-dimensional embeddings can provide strong predictive features without revealing which spectral, temporal, or management signals drive an estimate. Agricultural deployment therefore requires transparent reporting of model provenance, training data, preprocessing, uncertainty, and computational requirements. Addressing these limitations is necessary before broad claims of agricultural generality can be supported. The environmental and computational costs of large-scale pretraining also deserve explicit reporting because such requirements influence reproducibility, accessibility, and the feasibility of deployment by organizations with limited infrastructure [6]. Recent 2026 agricultural GFM evaluations identify transferability, interpretability, temporal sensitivity, and computational reproducibility as continuing limitations even when embeddings or multimodal models achieve strong task-level performance [6]. These limitations indicate that agricultural deployment requires evidence about where, when, and why a representation remains reliable.

Future Research Directions

Future research on geospatial foundation models (GFMs) for agricultural applications should prioritize agriculture-specific pretraining that captures crop calendars, phenological transitions, management events, and agronomic state variables rather than relying predominantly on generic image-reconstruction objectives. Training datasets should represent diverse cropping systems, smallholder regions, and under-sampled crops to reduce the risk of reproducing existing geographic and agricultural biases in pretrained representations [4]. Developing domain-specific pretraining objectives could therefore improve the ability of GFMs to recognize crop development stages, stress conditions, and management responses across heterogeneous agricultural landscapes. Another important direction is the development of physics-informed and hybrid foundation models that integrate learned geospatial representations with crop-growth and radiative-transfer models. Such integration could constrain model outputs according to physically plausible relationships among canopy structure, water availability, pigment dynamics, biomass accumulation, and yield formation [5]. Hybrid architectures may improve model extrapolation under environmental conditions that differ from the training distribution while also generating interpretable intermediate variables that can be related to established crop-growth processes.

Scale bridging also remains a critical research challenge because agricultural observations are acquired through UAVs, satellites, proximal sensors, and field-based measurements at substantially different spatial, temporal, and spectral scales. Future GFMs should jointly represent these observations while explicitly accounting for spatial resolution, spectral response, viewing geometry, and radiometric differences [2]. Recent scale-aware and wavelength-aware approaches provide potential architectural foundations for such systems [24]. A 2026 hierarchical foundation-model framework for agricultural parcel delineation further demonstrated how parameter-efficient adaptation, zero-shot segmentation, and multi-scale learning can support transfer across European and Chinese datasets [9]. These developments suggest that multi-scale foundation models could complement satellite-scale GFMs when agricultural management decisions require accurate field and parcel geometries.

The development of standardized agricultural benchmarks should accompany advances in model architecture. Future benchmarks should encompass multiple countries, crops, phenological stages, stress conditions, biophysical variables, and yield measurements while incorporating standardized spatial and temporal holdouts and multiple label budgets [5]. Existing resources, including GEO-Bench, PANGAEA, PhilEO, Crop Harvest, PASTIS, and Crop Net, provide important components but do not yet constitute a unified agriculture-centered evaluation framework [25]. Recent 2026 agricultural GFM benchmarking studies further indicate that model inputs, preprocessing procedures, and evaluation splits can substantially influence apparent transfer performance [1]. Consequently, future benchmark protocols should employ harmonized preprocessing and region-held-out evaluation so that reported performance more accurately reflects genuine geographic generalization.

How to cite this article: Wassaf A, Zeenat K. Geospatial Foundation Models for Intelligent Crop Monitoring: AI-Driven Multispectral Analysis, Yield Prediction, and Precision Agriculture. JOJ Horticulture & Arboriculture, 6(4). 555695. DOI: 10.19080/JOJHA.2026.06.555695

Uncertainty quantification and explainability should become integral components of agricultural GFM evaluation rather than optional additions. Future studies should report calibrated prediction intervals, error distributions across crops and regions, and explanations associated with meaningful spectral, temporal, and phenological features [2]. Yield-prediction research should additionally distinguish uncertainty originating from the model from uncertainty associated with ground-truth yield measurements [5]. Recent multimodal crop-stress research has demonstrated the value of evaluating uncertainty calibration under geographic and temporal distribution shifts [2]. Accordingly, future agricultural GFM studies should assess not only predictive accuracy but also whether uncertainty estimates remain reliable when models are applied to unfamiliar regions, seasons, and environmental conditions. Computational efficiency and edge deployment represent another important research direction. Although large foundation models can provide powerful representations, their computational requirements may limit adoption by regional agricultural agencies, research institutions, extension services, and farm-level decision-support systems. Future research should therefore investigate knowledge distillation, model compression, parameter-efficient adaptation, and lightweight architectures capable of operating with limited computational resources [28]. Approaches such as Presto demonstrate the potential for extracting useful agricultural representations without relying on the computational footprint of the largest foundation models.

The integration of language-based and agentic interfaces could further expand the practical accessibility of agricultural GFMs. Future systems could allow agronomists, extension specialists, and agricultural advisors to query geospatial representations through structured natural-language questions concerning crop stress, phenology, spatial variability, or expected yield. However, such systems should maintain traceability to the underlying imagery, weather observations, field measurements, and uncertainty estimates [2]. Their evaluation should therefore emphasize factual reliability, evidence traceability, and decision usefulness rather than interface novelty alone. Ultimately, farm-level validation is necessary to determine whether improvements in foundation-model representations translate into measurable agricultural benefits. Future studies should employ randomized, matched, or stepped-wedge on-farm evaluations to quantify changes in realized yield, input use, production costs, and environmental outcomes [5]. Recent 2026 studies involving agricultural field boundaries and foundation models demonstrate that these approaches can be evaluated in fragmented smallholder landscapes, while also showing that zero-shot performance and transferability remain sensitive to spatial scale and input configuration [3].

Consequently, field trials should evaluate not only representation quality but also the practical consequences of model uncertainty for agricultural decisions [10]. To improve comparability and reproducibility across studies, future research should also adopt a minimum reporting framework covering the principal sources of methodological ambiguity identified in the literature [6]. Each study should report the exact model's name, version, parameter count, checkpoint, and fine-tuning strategy; the sensor and product level, spectral bands, preprocessing, cloud masking, compositing window, and number of observations; and the source, collection date, spatial support, noise characteristics, and distribution of labels. Evaluation protocols should clearly describe spatial blocking, held-out years or regions, and whether pretraining scenes overlap with test areas. Studies should additionally compare against from-scratch, index-based, regional baseline, and strong task-specific models, while reporting performance across multiple training-set sizes and repeated runs. Finally, reporting should include uncertainty calibration, computational cost, and decision relevance. Researchers should provide prediction-interval coverage or calibration metrics, error stratification by crop, region, and season, as well as pretraining and adaptation computational requirements, inference time per hectare, and hardware specifications. For practical agricultural applications, studies should also report forecast lead time, spatial resolution relative to actual management units, and evidence of realized agronomic or economic benefits. Such standardized reporting would strengthen reproducibility, enable meaningful comparison between GFMs and conventional task-specific models, and help determine whether advances in geospatial representation learning ultimately translate into measurable improvements in crop monitoring and yield enhancement.

Conclusions

Geospatial foundation models offer a promising route for multispectral crop monitoring because these models can decouple representation learning from the limited labelled data available for many agricultural tasks. The strongest current evidence concerns crop and cropland mapping, especially in settings where pretrained representations can be adapted with relatively little supervision. However, benchmark results also show that GFMs do not uniformly outperform strong supervised baselines, and the evidence base for direct yield estimation, pre-visual stress detection, and realized farm-level benefits remains comparatively limited. Such evidence would establish when geospatial foundation models provide a scientifically and operationally meaningful advantage for crop monitoring. The evidence reviewed here indicates that the most defensible role of current GFMs is as reusable representation infrastructure whose agricultural value must be demonstrated task by task. Crop mapping currently provides clearer evidence than direct yield enhancement, while stress detection, biophysical retrieval, and intervention outcomes remain less comprehensively evaluated. Model flexibility, label efficiency, computational cost, and geographic transfer should therefore be considered jointly rather than inferred from a single benchmark metric.

The next stage of progress should priorities evidence and agricultural relevance rather than model scale alone. Key requirements are agriculture-specific and geographically balanced pretraining, standardised spatially and temporally honest

How to cite this article: Wassaf A, Zeenat K. Geospatial Foundation Models for Intelligent Crop Monitoring: AI-Driven Multispectral Analysis, Yield Prediction, and Precision Agriculture. JOJ Horticulture & Arboriculture, 6(4). 555695. DOI: 10.19080/JOJHA.2026.06.555695

benchmarks, explicit UAV–satellite scale bridging, multimodal and physics-informed models, calibrated uncertainty, and farm-level intervention studies. Recent 2026 agricultural benchmarking of Alpha Earth embeddings further illustrates both the potential of foundation representations and persistent limitations in transferability, interpretability, and temporal sensitivity. Future progress should therefore be measured through geographically independent agricultural benchmarks and field-level outcome studies rather than model size or a single aggregate metric. Such evidence would establish when geospatial foundation models provide a scientifically and operationally meaningful advantage for crop monitoring. The evidence reviewed here indicates that the most defensible role of current GFMs is as reusable representation infrastructure whose agricultural value must be demonstrated task by task. If these elements are developed together, GFMs can be evaluated not only as reusable representation technologies but as components of measurable, transparent, and decision-relevant agricultural systems.

References

1. Osei M (2023) Remote Sensing in Field Crop Monitoring: A Comprehensive Review of Sensor Systems, Data Analyses and Recent Advances. Remote Sensing 15: 354.
2. Kganyago M, Adjorlolo C, Mhangara P, Soelen LT (2024) Optical Remote Sensing of Crop Biophysical and Biochemical Parameters: An Overview of Advances in Sensor Technologies and Machine Learning Algorithms for Precision Agriculture. Computers and Electronics in Agriculture 218: 108730.
3. Wang Y, Albrecht CM, Ait Ali Braham N, Mou L (2022) Self-Supervised Learning in Remote Sensing: A Review. IEEE Geoscience and Remote Sensing Magazine 10(4): 213-247.
4. Jakubik J (2023) Foundation Models for Generalist Geospatial Artificial Intelligence arXiv.
5. Zhang F (2026) A Review of Crop Yield Estimation on Pixel and Field Scales from Remotely Sensed Data. Science of Remote Sensing 13: 100342.
6. Ma Y, Shen Y, Swatantra A, Lobell DB (2026) Harvesting Alpha Earth Benchmarking the Geospatial Foundation Model for Agricultural Downstream Tasks. International Journal of Applied Earth Observation and Geoinformation 149: 105258.
7. Oikonomidis AC, Cathal A, Kassahun (2022) Deep Learning for Crop Yield Prediction: A Systematic Literature Review. New Zealand Journal of Crop and Horticultural Science 51(1): 1-26.
8. Segarra J, Araus JL, Kefauver SC (2022) Farming and Earth Observation: Sentinel-2 Data to Estimate Within-Field Wheat Grain Yield. International Journal of Applied Earth Observation and Geoinformation 107(1): 102697.
9. Tseng G, Zvonkov IC, Nakalembe, H Kerner (2021) Crop Harvest: A Global Dataset for Crop-Type Classification. NeurIPS Datasets and Benchmarks Track.
10. Priyabrata K, Shyh W, Manzur (2024) Crop Monitoring by Multimodal Remote Sensing: A Review. Remote Sensing Applications: Society and Environment 33(10): 101093.
11. Li F (2026) Earth Observation Derived Yield Forecasting and Estimation in Low- and Lower-Middle-Income Countries Dominated by Smallholder Agriculture: A Review. International Journal of Applied Earth Observation and Geoinformation 148: 105216.
12. Demissie WA, Sebastiani L, Rossetto R (2026) Integration of Artificial Intelligence and Remote Sensing for Crop Yield Prediction and Crop Growth Parameter Estimation in Mediterranean Agroecosystems: Methodologies, Emerging Technologies, Research Gaps, and Future Directions. European Journal of Agronomy 173: 127894.
13. Ganie MW, Wani MA, Wani AH (2026) Machine Learning Approaches for Crop Yield Prediction: A Systematic Literature Review. Remote Sensing Applications: Society and Environment 1339-1350.
14. Yu Wet (2026) Rice Yield Prediction in Unseen Years at Field Level with High-Resolution Gross Primary Productivity Derived from Sentinel-2 Imagery. Remote Sensing of Environment 16(3):10-3390.
15. Cong YS, Khanna C, Meng P, Liu E, Rozi, et al. (2022) Sat MAE: Pre-Training Transformers for Temporal and Multi-Spectral Satellite Imagery. Advances in Neural Information Processing Systems 35:197-211.
16. Fuller A, Millard K, JR Green (2023) CROMA: Remote Sensing Representations with Contrastive Radar-Optical Masked Autoencoders. Advances in Neural Information Processing Systems 36: 5506-5538.
17. Szwarc Man D (2024) Prithvi-EO-2.0: A Versatile Multi-Temporal Foundation Model for Earth Observation Applications arXiv.
18. Xiong Z (2024) Neural Plasticity-Inspired Multimodal Foundation Model for Earth Observation arXiv.
19. Clay Foundation (2024) Clay Foundation Model: An Open-Source AI Model for Earth Observation. Model documentation.
20. Tseng G (2025) Galileo Learning Global & Local Features of Many Remote Sensing Modalities. Proceedings of the 42nd International Conference on Machine Learning 267: 60280-60300.
21. Hong D (2024) Spectral Spectral Remote Sensing Foundation Model. IEEE Transactions on Pattern Analysis and Machine Intelligence 46(8): 5227-5244.
22. Bastani F, Wolters P, Gupta J Ferdinando A Kembali (2023) SatlasPretrain: A Large-Scale Dataset for Remote Sensing Image Understanding. Proceedings of the IEEE/CVF International Conference on Computer Vision pp. 16772-16782.
23. Astruc G, Gonthier NC, Mallet, L Landrieu (2025) Any Sat One Earth Observation Model for Many Resolutions Scales and Modalities. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition pp. 19530-19540.
24. Lacoste A (2023) GEO-Bench Toward Foundation Models for Earth Monitoring. Advances in Neural Information Processing Systems 36.
25. Fiber C, Camilleri L, Luyts A, Dionelis N, Le Saux B (2024) PhilEO Bench: Evaluating Geo-Spatial Foundation Models. IEEE International Geoscience and Remote Sensing Symposium (IGARSS).
26. Marsocci V (2024) PANGAEA: A Global and Inclusive Benchmark for Geospatial Foundation Models arXiv.
27. Page MJ (2021) The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews BMJ 372: n71.
28. Xiao A (2024) Foundation Models for Remote Sensing and Earth Observation: A Survey arXiv.
29. Lu B, Proctor C, He Y (2021) Investigating Different Versions of PROSPECT and PROSAIL for Estimating Spectral and Biophysical Properties of Photosynthetic and Non-Photosynthetic Vegetation in Mixed Grasslands. GI Science Remote Sensing 58(7): 1-18

How to cite this article: Wassaf A, Zeenat K. Geospatial Foundation Models for Intelligent Crop Monitoring: AI-Driven Multispectral Analysis, Yield Prediction, and Precision Agriculture. JOJ Horticulture & Arboriculture, 6(4). 555695. DOI: 10.19080/JOJHA.2026.06.555695

30. Xu H (2022) Optical Vegetation Indices for Monitoring Terrestrial Ecosystems Globally. Nature Reviews Earth & Environment 3(7): 477-493.

31. Vidican R (2023) Using Remote Sensing Vegetation Indices for the Discrimination and Monitoring of Agricultural Crops: A Critical Review. Agronomy 13(12): 3040.

32. Duan B, Fang S, Gong Y, Peng Y, Wu X, Zhu R (2021) Remote Estimation of Grain Yield Based on UAV Data in Different Rice Cultivars under Contrasting Climatic Zone. Field Crops Research 267(11): 108148.

33. Marshall M, (2022) Field-Level Crop Yield Estimation with PRISMA and Sentinel-2. ISPRS Journal of Photogrammetry and Remote Sensing 187(2): 191-210.

34. Ju J (2025) The Harmonized Landsat and Sentinel-2 Version 2.0 Surface Reflectance Dataset. Remote Sensing of Environment 324: 114723.

35. Reed CJ (2023) Scale-MAE A Scale-Aware Masked Autoencoder for Multiscale Geospatial Representation Learning. Proceedings of the IEEE CVF International Conference on Computer Vision pp. 4088-4099.

36. Sun X (2022) Ring Mo A Remote Sensing Foundation Model With Masked Image Modeling. IEEE Transactions on Geoscience and Remote Sensing 61: 1-22.

37. Li W (2026) Agri FM A Multi-Source Temporal Remote Sensing Foundation Model for Agriculture Mapping. Remote Sensing of Environment 334: 115234.

38. Han B, Zhang S, Shi X, Reichstein M (2024) Bridging Remote Sensors with Multisensory Geospatial Foundation Models. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition pp. 27852-27862.

39. Suresh A (2025) Fine-Tuning the Prithvi Foundation Model for Crop-Type and Health Mapping in Smallholder Farms. Procedia Computer Science 270: 5055-5064.

40. Yan J, Gu X, Chen Y (2025) CropSTS A Remote Sensing Foundation Model for Cropland Classification with Decoupled Spatiotemporal Attention. Remote Sensing 17(14): 2481.

41. Xue S (2026) Crop Classification Mapping in Multi-Cloud Regions An Integrated Approach Using GF-SG Time Series Reconstruction and 3D Deep Convolutional Networks. Science of Remote Sensing 13: 100361.

42. Halder K (2026) High-Resolution Maize Yield Mapping Across Africa Using Earth Observation and Machine Learning, Deep Learning, and Foundation Model. Science of Remote Sensing 13: 100344.

43. Ye F (2026) A Review of Remote Sensing-Crop Model Data Assimilation Addressing Challenges in Data-Scarce and Spatially Heterogeneous Regions. European Journal of Agronomy.

44. Gamewell Y, Chaurasia MA (2026) Explainable Multi Season Spartoi Temporal Deep Learning Framework for Crop Yield Forecasting Using Sentinel-2 Remote Sensing Data. Discover Computing 29: 452.

45. Piyare NS, Baulkani S (2026) An Explainable Spartoi-Temporal Deep Learning Framework for Crop Yield Prediction and Recommendation. Scientific Reports 16(1): 24662.

46. Jing K, Zhang M, Wu B, Xie Y, Li M (2026) Distilling Geospatial Foundation Model Semantics into Interpretable Sentinel-1/2 Proxies for Few-Shot and Cross-Year Crop Type Classification. Big Earth Data.

47. Wang W, Guo H, He S, Qi F, Samat A, et al. (2026) AI-Driven Precision Mapping of Tea Plantations Using Alpha Earth Foundations Scalable Solution for Smart Agricultural Monitoring. Agriculture 16(4): 412.

48. Gernot VS, Landrieu FL (2021) Panoptic Segmentation of Satellite Image Time Series with Convolutional Temporal Attention Networks. Proceedings of the IEEE/CVF International Conference on Computer Vision pp. 4872-4881.

49. Li W (2026) Agri FM A Multi-Source Temporal Remote Sensing Foundation Model for Agriculture Mapping. Remote Sensing of Environment 334: 115234.

50. Tripathy PK, Baylis K, Wu J, Watson Jiang R (2026) Zero-Shot Inference Strategies for Smallholder (<0.1 ha) Agriculture Field Delineation with the Segment Anything Foundation Model. Science of Remote Sensing 13: 100425.

51. Zhu Y, Pan Y, Hu T, Liu Y, Wang Y (2026) A Hierarchical Foundation Model Framework for Generalizable Agricultural and Crop Parcel Delineation. Computers and Electronics in Agriculture 245: 111523.

This work is licensed under Creative Commons Attribution 4.0 License
DOI: 10.19080/JOJHA.2026.06.555695

Your next submission with Juniper Publishers will reach you the below assets

- Quality Editorial service
- Swift Peer Review
- Reprints availability
- E-prints Service
- Manuscript Podcast for convenient understanding
- Global attainment for your research
- Manuscript accessibility in different formats
  (Pdf, E-pub, Full Text, Audio)
- Unceasing customer service

Track the below URL for one-step submission
https://juniperpublishers.com/online-submission.php

How to cite this article: Wassaf A, Zeenat K. Geospatial Foundation Models for Intelligent Crop Monitoring: AI-Driven Multispectral Analysis, Yield Prediction, and Precision Agriculture. JOJ Horticulture & Arboriculture, 6(4). 555695. DOI: 10.19080/JOJHA.2026.06.555695