Review

Not peer-reviewed version

AlphaEarth Foundations (AEF) in Earth Observation: A Systematic Review of Applications and Practices

Shuyang Hou , Haoyue Jiao , Ziqi Liu , Lutong Xie , Guanyu Chen , Shaowen Wu , Zhangyan Xu ,
Zengjie Wang , Shaoqing Tang , Yaxian Qing , Jianyuan Liang , Xuefeng Guan , Huayi Wu *

Posted Date: 14 May 2026

doi: 10.20944/preprints202605.0981.v1

Keywords: AlphaEarth Foundations; surface embedding; unified representation; Earth observation; remote sensing review

doi:10.20944/preprints202605.0981.v1

Review

AlphaEarth Foundations (AEF) in Earth Observation: A Systematic Review of Applications and Practices

Shuyang Hou [1,†], Haoyue Jiao [1,2,†], Ziqi Liu [1], Lutong Xie [1], Guanyu Chen [1,2], Shaowen Wu [1], Zhangyan Xu [1], Zengjie Wang [1,2], Shaoqing Tang [1], Yaxian Qing [1], Jianyuan Liang [1], Xuefeng Guan [1,*] and Huayi Wu [1,*]

[1] State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan, China
[2] School of Resource and Environmental Sciences, Wuhan University, Wuhan, China
* Correspondence: guanxuefeng@whu.edu.cn (X.G.); wuhuayi@whu.edu.cn (H.W.)
† These authors contributed equally to this work. .

Abstract

AlphaEarth Foundations (AEF) unifies multi-modal and multi-temporal observations into analysis-ready low-dimensional representations, introducing a representation-driven paradigm that addresses key limitations of traditional task-centric approaches, including high engineering complexity, limited cross-regional transferability, and strong dependence on labeled data. This paper presents the first systematic review of AEF, synthesizing 23 research articles published up to March 2026 from three perspectives: conceptual positioning, technical framework, and application practices. It first clarifies the distinctions between AEF and related paradigms, including foundation models, remote sensing large models, and existing unified embedding approaches. It then summarizes the organization of its data system and key technical components. Based on a systematic literature survey, the paper further provides a structured synthesis of current studies in terms of thematic and regional distribution, scientific questions, usage patterns, and evaluation methods. The analysis indicates that AEF represents an important step toward a paradigm shift from task-driven to representation-driven Earth observation, offering clear advantages in reducing engineering barriers, enabling cross-regional transfer, and establishing a unified environmental semantic foundation. However, limitations remain in temporal resolution, semantic interpretability, and performance in specific task scenarios. Future work should focus on enhancing dynamic representation capabilities, cross-domain adaptation mechanisms, multi-source integration frameworks, systematic evaluation and annotation systems, and asset-oriented applications. By delineating the capability boundaries and applicable contexts of AEF from both methodological and empirical perspectives, this study provides a systematic reference for the standardized development and rational application of unified surface embeddings as an emerging geospatial information infrastructure.

Keywords: AlphaEarth Foundations; surface embedding; unified representation; Earth observation; remote sensing review

---

1. Introduction

Earth Observation (EO) refers to the technical system for continuously and systematically acquiring and quantitatively characterizing the state and dynamic changes of the Earth's surface through satellites, aerial platforms, and ground sensors (Balsamo et al., 2018; Li et al., 2012). As a critical infrastructure for global environmental perception, EO provides essential data support for major issues such as the assessment of United Nations Sustainable Development Goals, food security monitoring, ecosystem dynamics analysis, and carbon sink accounting (Anderson et al., 2017; Kavvada et al., 2020; Wu et al., 2023). Observational data with high spatial and temporal resolution, long time series, and cross-regional comparability form the fundamental basis for surface information extraction, change understanding, and mechanism modeling (Wu et al., 2025; Zhu & Woodcock, 2012).

In recent years, the rapid growth in the number of Earth observation satellites, along with the continuous deployment of multi-type sensors such as optical, synthetic aperture radar (SAR), and hyperspectral instruments, has led to ongoing expansion in the spatial resolution, temporal density, and observational dimensions of EO data (Yao et al., 2019). Although multi-source EO data provide an unprecedented observational foundation for modeling surface processes, practical applications still face significant challenges in achieving stable, continuous, and transferable surface state characterization: ultra-large-scale data significantly increase costs for data screening, storage, and computation (Koldasbayeva et al., 2024; Xu et al., 2022); non-uniform temporal sampling and irregular observation intervals add uncertainty to change modeling and attribution analysis (Constantin et al., 2022); inconsistent spatial resolution and observation scales limit cross-scale alignment, knowledge transfer, and regional generalization capabilities (Claverie et al., 2018; Kartal et al., 2018).

The development of deep learning has partially alleviated the reliance on manual feature design in traditional remote sensing interpretation and enhanced the representation capability of complex surface patterns through end-to-end learning (Ma et al., 2019; Zhao et al., 2026). However, its mainstream paradigm remains task-centric, with model performance heavily dependent on large-scale, high-quality annotations and stable training distributions, leading to significant degradation in cross-regional, cross-temporal, and cross-sensor scenarios (Al-Emadi et al., 2025; J. Zeng et al., 2024; W. Zeng et al., 2024; Zhu et al., 2023). Simultaneously, multi-source heterogeneous observations and non-uniform spatiotemporal sampling continually challenge the implicit assumption of “fixed grid—fixed time step—identically distributed training,” making it difficult for models to generalize consistently (Constantin et al., 2022; Koldasbayeva et al., 2024; Zhou et al., 2023). Moreover, due to sensor imaging characteristics and observational conditions, EO data still require complex preprocessing steps such as radiometric calibration, geometric correction, cloud shadow removal, and cross-sensor harmonization before application, further increasing the engineering barriers for model training and deployment (Ju et al., 2025; Reinartz et al., 2011; Vermote et al., 2016). Consequently, the scarcity of high-quality annotated data and the high engineering costs jointly constrain the sustainability of task-centric supervised learning paradigms at a global scale.

Against the backdrop of non-negligible computational costs and engineering complexity, where deep learning models have not yet become low-barrier infrastructure resources, relying solely on repetitive training of larger-scale models cannot fundamentally improve the cost-effectiveness and scalability of EO applications (Haut et al., 2021; Persello et al., 2022). The current key issue is no longer merely model architecture design, but rather how to construct a general-purpose representation layer for surface states: one that assimilates multi-source, multi-temporal observations into a unified, compact, and transferable representation, enabling downstream mapping, monitoring, and estimation tasks to be efficiently completed with minimal or even no additional annotations (Jin et al., 2026; Pascual & Guerra-Hernández, 2026).

Based on this concept, AlphaEarth Foundations (AEF), proposed by Google DeepMind, attempts to address the long-standing bottlenecks in data processing and representation within the EO domain in the form of “analysis-ready representation data products” (Brown et al., 2025). AEF integrates multi-source EO imagery globally, including optical, multi-temporal SAR, and LiDAR data, combining geographic location, temporal context, and related semantic information. Through a unified representation learning framework, it maps observations from different sensors, modalities, and acquisition times into consistent surface embeddings. Its output employs a globally uniform spatial grid, generating 64-dimensional low-dimensional embedding vectors for each surface unit at approximately 10 m spatial resolution, and forming annual embedding layers through temporal aggregation. This representation encapsulates key information from multi-source observations in a compact and well-aligned numerical format, significantly reducing the engineering burden associated with raw image preprocessing, cross-sensor alignment, and feature construction, allowing downstream applications to directly perform mapping and analysis based on the embeddings. Existing studies have reported that traditional shallow supervised learning models built upon AEF embeddings can achieve performance comparable to, and in some cases even more stable than, deep learning methods across various surface mapping tasks (Fang et al., 2026; Ma et al., 2026; Ryan et al., 2026). AEF has publicly released global annual embedding data covering 2017–2024, providing a new data foundation for cross-modal and cross-temporal consistency analysis.

However, existing literature lacks a systematic review of the applicability boundaries and methodological assumptions of AEF. As AEF is increasingly applied to diverse tasks and scenarios, some studies have reported situations where its unified embeddings exhibit limited applicability or unstable performance under specific conditions; however, these discussions are often scattered across individual application works and have not yet formed a clear general understanding. Key issues identified in current research mainly include: differences in the applicability of unified embeddings under varying task types, spatial scales, and cross-regional distribution transfers; potential degradation in representational capacity under conditions of temporal gaps, inconsistent spatial resolutions, and multi-source heterogeneous observations; and the relationship between the unified embedding layer and physical consistency constraints, uncertainty descriptions, and error propagation analysis when used as an intermediate representation. These issues directly affect the usage mode and developmental positioning of AEF in practical applications, i.e., whether it is better suited as an engineered data product under specific conditions or as a general-purpose geospatial information infrastructure with further expansion potential. Given that current related studies primarily focus on specific application demonstrations and empirical usage reports, there is a need for a systematic synthesis and summary of AEF's methodological framework, design trade-offs, applicable scenarios, and existing insights to guide future research and standardized applications.

Therefore, this paper is not a simple restatement of the original AEF paper, but rather a distillation and summarization centered on its core ideas and technical features, aiming to help readers develop an overall understanding of the methodology. Based on this, the paper systematically reviews current research surrounding AEF, summarizes its main application directions and existing insights, and analyzes its advantages and potential limitations in light of methodological design and application feedback, while also forecasting possible future developments. The article is structured as follows: Section 2 distinguishes AEF from related concepts such as "foundation models" and "remote sensing large models," clarifying its methodological positioning within the development trajectory of unified embedding approaches; Section 3 outlines the technical framework and key design elements of AEF, explaining its data organization and technical highlights; Section 4 reviews existing studies, introduces the scope of literature retrieval, and categorizes relevant empirical work in terms of research themes, regional distribution, core scientific questions, usage modes, and evaluation methods; Section 5 synthesizes the strengths and limitations of AEF based on technical features and application evidence, analyzes the interpretability exploration paradigms across major dimensions, and offers perspectives on future development directions. An overview of the review structure is presented in Figure 1.

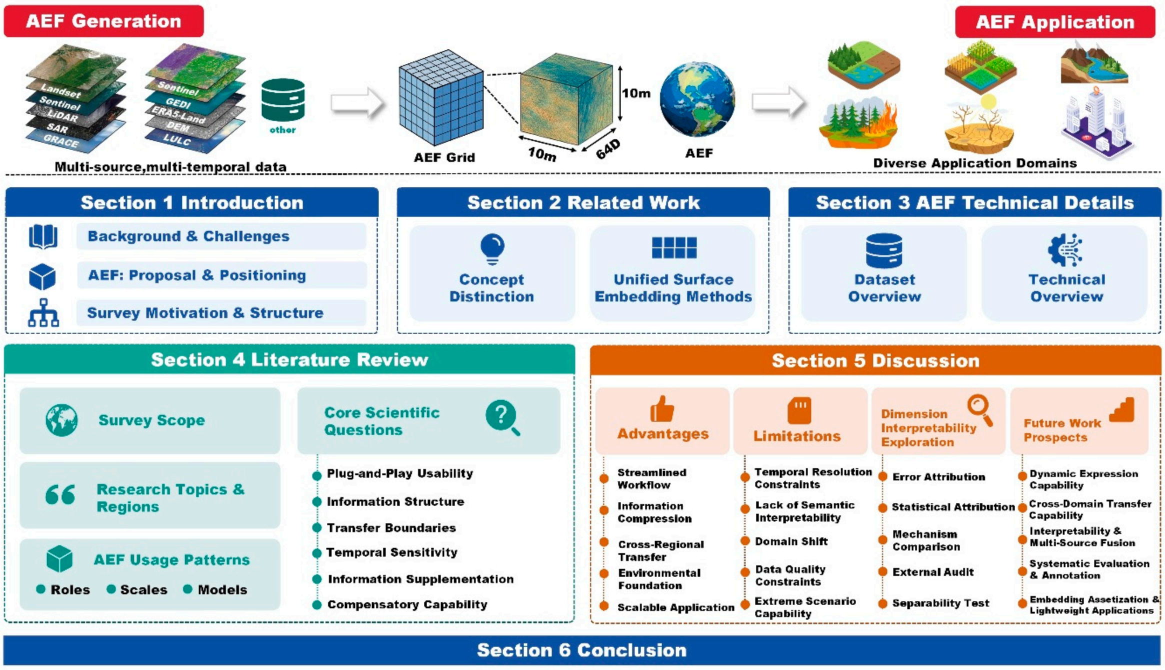

Figure 1. Overall framework of the review structure.

2.2. Related Work

2.2.1. Concept Distinction

It is necessary to distinguish AEF from the frequently mentioned concepts such as "foundation models" and "remote sensing large models" in the current EO field. In the vision domain, foundation models typically refer to parameterized general-purpose models trained on large-scale data, such as CLIP (Gao et al., 2024; Hafner et al., 2021), DINO (Yamamoto et al., 2025; Zhang et al., 2022), and Segment Anything (Kirillov et al., 2023). Their core characteristic lies in the fact that model parameters themselves serve as transferable knowledge carriers, requiring users to invoke their capabilities through model inference or downstream fine-tuning. In recent years, the "remote sensing large models" proposed in the remote sensing field have largely followed this paradigm, emphasizing cross-task pre-training, model parameter scale expansion, and fine-tuning and transfer capabilities for specific applications, such as SkySense (Guo et al., 2024) and Prithvi (Szwarcman et al., 2025).

In contrast to the aforementioned model paradigms, AEF is not released in the form of a parameterized model that can be directly invoked or fine-tuned, but rather as a surface embedding representation data product. The methodology of AEF is publicly available through academic papers, but its specific training process and model parameters are not disclosed. What users actually interact with and use are the embedding results generated under unified spatial and temporal scales (Brown et al., 2025). Therefore, AEF differs significantly from typical parameterized foundation models or remote sensing large models, and is more appropriately understood as a unified representation framework for EO, along with the analysis-ready embedding data products generated by this framework. For parameterized foundation models and remote sensing large models, several comprehensive review works have systematically summarized aspects such as model architecture, training paradigms, and application performance (Huang et al., 2025; Huo et al., 2025; Lu et al., 2025; Xiao et al., 2025). In comparison, systematic reviews focusing on representation-type data products published in the form of embeddings within the EO field, especially those targeting AEF and its embedding products, remain relatively scarce. The methodological characteristics, applicable scenarios, and relationships with existing model paradigms still require further clarification and summarization.

2.2.2. Unified Surface Embedding Method

AEF is not the origin of the concept of unified surface embedding. Prior to its proposal, the remote sensing field had already seen a series of explorations aimed at reducing the reliance on raw image processing and manual labeling for downstream tasks by learning compact representations. These studies typically focused on specific observational elements, application themes, or regional scales, such as climate variable inversion (Wang et al., 2021), land cover classification (Pelletier et al., 2019), and ecological condition monitoring (Gong et al., 2025; Solórzano et al., 2023). Early representative works include Tile2Vec, a self-supervised remote sensing embedding method based on spatial neighborhood consistency (Jean et al., 2019), and SeCo, which learns transferable remote sensing feature representations by introducing temporal and seasonal consistency constraints (Manas et al., 2021). In addition, contrastive learning studies conducted using Sentinel-2 and Landsat data have explicitly aimed to obtain low-dimensional, generalizable remote sensing embeddings, attempting to reduce dependence on manually designed features and dense annotations (Liu et al., 2024; Marsocci & Audebert, 2025). At the same time, research on temporal sequence representation of multi-temporal remote sensing images has begun to explore the use of temporal coherence or sequence modeling approaches to learn surface state representations, laying the foundation for subsequent temporal embedding methods (Cong et al., 2022; Dong et al., 2020; Xu et al., 2024). These early efforts have already demonstrated the core idea of replacing manual feature design with representation learning; however, their training and validation often rely on specific datasets, regions, or task scenarios, thus exhibiting certain limitations in spatial scalability, cross-regional consistency, and long-term stable usage, making it difficult to support unified global applications.

With the continuous accumulation of multi-source remote sensing data and the development of self-supervised learning methods, the research focus has gradually shifted from “task-specific features” to “universal surface representations”. Related work has started to introduce constraints such as temporal consistency, cross-modal alignment, and contrastive learning to learn embedding representations that support multi-task transfer from multi-source remote sensing time series. For example, Presto constructs reusable surface representations through self-supervised training on multi-source remote sensing time series (Tseng et al., 2023), while methods like TESSERA further explore cross-sensor feature alignment and temporal consistency modeling, demonstrating certain transfer potential at the regional scale or within specific task scenarios (Feng et al., 2025). These studies collectively reflect a methodological shift from “model-driven tasks” to “representation-supported analysis”.

However, the aforementioned embedding methods are often still implemented as models or feature extractors, requiring users to complete model inference, data alignment, and deployment independently during practical applications, resulting in relatively complex overall engineering processes, and making it difficult to directly reuse or compare representations across different studies. AEF builds upon the concept of unified embedding and further advances this paradigm toward a global scale and product level by integrating multi-modal EO data and releasing annual embedding layer data at fixed time intervals, significantly reducing the complexity of data processing and model usage on the user side. Moreover, AEF emphasizes automatically identifying discriminative features for characterizing surface states through the learning process, without relying on high-quality, complete labeled data across all time phases. Instead, it extracts stable representations from incomplete and noisy observations via self-supervision and data consistency constraints. This data-driven approach to representation design enables AEF to potentially generalize across regions and time periods without the need for manually defined key features. Due to the comprehensive trade-offs made in terms of scale, release format, and learning paradigm, AEF exhibits distinct stage characteristics within the existing trajectory of unified embedding research, warranting a systematic review and summary of its conceptual origins, technical evolution, methodological features, and applications.

doi:10.20944/preprints202605.0981.v1

3. AEF Technical Details

3.1. Dataset Overview

As a globally covering “super map”-style embedded field model, AEF’s core capabilities are closely tied to its data support. Clarifying its data composition is a critical prerequisite for defining the model’s capability boundaries, applicable scenarios, and performance limits.

The AEF data system can be clearly divided into three categories based on function: training input, training target, and evaluation data (Brown et al., 2025). Among these, training input serves as the foundational data source for model learning, primarily including multi-source remote sensing observations such as Sentinel-2, Sentinel-1, and Landsat 8/9. These provide multimodal, multitemporal surface information to support feature extraction and spatiotemporal representation learning. Training targets guide the model’s learning direction and include GEDI vegetation height, ERA5-Land climate variables, GRACE terrestrial water storage, DEM topographic information, and some land cover and textual semantic data. These are used to guide the model in learning surface properties and associations between multiple sources through cross-source reconstruction, contrastive learning, or weakly supervised constraints. Evaluation data are further categorized into classification, regression, and change detection, including LCMAP land cover, OpenET evapotranspiration, and ASTER GED emissivity, among others, to test the model’s generalization ability under sparse labeling conditions. It should be noted that some data sources serve dual roles as both inputs and targets. For example, Sentinel-2, Landsat 8/9, and Sentinel-1 not only act as training inputs providing basic observational information but also participate as training targets in cross-source reconstruction tasks to enhance the model’s ability to depict surface details and multisource consistency. In contrast, data such as GRACE, GEDI, and NLCD, which have coarser spatial resolution, limited spatial coverage, or sparse sampling, are more suitable as supervisory signals or semantic constraints rather than as globally unified inputs. On the other hand, textual targets (e.g., Wikipedia geographic entries and GBIF species records) are mainly used for semantic alignment, while image-based targets primarily serve cross-source reconstruction; together, they promote the formation of a unified representation at both the physical attribute and high-level semantic levels within the model. The specific data composition is shown in Table 1.

Table 1. Datasets used by AEF and their functional categorization.

Data Category	Data Source Name	Product/Data Set Version	Key Parameters	Purpose	Notes
Optical data	Sentinel-2	Copernicus Sentinel-2 L1C	10–20 m, multispectral, revisit period of approximately 5 days	Training input, training target	Provides high-resolution spectral information, using cloud masking and dynamic range compression to enhance data stability
	Landsat 8/9	Landsat 8/9 C2 T1 TOA	15–100 m, including thermal infrared	Training input, training target	Provides long-time series and thermal information supplementation, enhancing cross-sensor consistency

Radar data	Sentinel-1	Sentinel-1 GRD	10 m, VV/VH/HH/HV	Training input, training target	Provides all-weather structural and roughness information, and enhances robustness through noise clipping and artifact simulation
	ALOS PALSAR-2	Level 2.2 ScanSAR	25m, HH/HV	Training target	Provides long-wavelength penetration vegetation structure information, but is used only as a supervisory signal due to discontinuous coverage
LiDAR data	GEDI	GEDI L2A Canopy Height	25 m, canopy height	Training target	Provides sparse but highly reliable three-dimensional structural reference
Climate/Gravity data	ERA5-Land	Monthly Aggregated	~11 km, meteorological variables	Training target	Provide climatic background constraints to help distinguish environmental drivers from surface structure differences
	GRACE/GRACE-FO	Mass Grids V03	~55km, water storage	Training target	Provide cross-scale hydrological change signals and participate in training with lower weights
Topographic data	Copernicus DEM	GLO-30 DSM	30m, elevation, slope, aspect	Training target	Provide static topographic constraints to assist in characterizing hydrological and vegetation environments
Land cover data	NLCD	NLCD 2019/2021	30 m, 16 land cover classes	Training target,	Provide human-annotated semantic

				evaluation	references, mainly used for weak supervision and evaluation within the United States
Text/species data	Wikipedia	Snapshot as of 2024-04-21	≥100-word English text, including geographic coordinates	Training target	Provide geographic and ecological semantic descriptions for text alignment learning
	GBIF	2017–2023 species records	Taxonomic resolution down to species level, spatial accuracy ≤240 m	Training target, evaluation	Provide biodiversity semantic anchors, limiting the number of samples per category to avoid common species dominating the training
Evaluation-specific data	LCMAP	USGS LCMAP CONUS	30 m, land cover and change	Evaluation	Used for validating classification and change detection performance
	OpenET Ensemble	OpenET v2.0	30 m, monthly evapotranspiration	Evaluation	Used to test continuous variable regression task performance
	ASTER GED	AG100 V003	100 m, 8.3 μm emissivity	Evaluation	Used to test thermal property representation and cross-temporal generalization ability

3.2. Technical Overview

After clarifying the data framework of AEF, the next critical question is: how can these highly heterogeneous observations, characterized by uneven spatiotemporal sampling, variable quality, and significant scale differences, be organized into a unified representation system that is trainable, stable in output, and reusable on a global scale? Unlike most remote sensing self-supervised methods designed for single sensors or specific tasks, AEF targets a multi-source, multi-modal observation set that simultaneously integrates optical, SAR, LiDAR, climatic, and textual semantic information (Brown et al., 2025). Under this setting, technical challenges extend beyond network architecture design and are primarily concentrated in three key aspects: data alignment and quality control, training objective and constraint design, and the construction and deployment form of the representation space. Without proper data preprocessing and observation window organization, multi-source information cannot effectively collaborate within a unified spatiotemporal semantic coordinate system; if the training phase fails to suppress "cross-source shortcuts" and distribution bias, the model may learn statistical features unrelated to surface states, thereby losing stability under cross-regional and temporal conditions; and if the representation formation stage lacks a unified, compact, and comparable latent space design, even effective training results will struggle to produce reusable, scalable analysis-ready data products.

Therefore, this section outlines the technical framework of AEF around its key design trade-offs: first summarizing the unified processing strategies and spatiotemporal organization methods applied to multi-source data before model input; secondly, generalizing its training mechanism centered on cross-source reconstruction and consistency constraints, explaining how it forms transferable representations under weak supervision; finally, discussing how representation formation strategies such as spherical embedding and low-dimensional compression serve storage efficiency, similarity retrieval, and downstream applications without fine-tuning. To facilitate an overall understanding, Table 2 summarizes the core functions, implementation approaches, and potential limitations of each technical module in AEF, laying the foundation for subsequent application evaluation and limitation discussions.

Table 2. Main technical modules, functions, and limitations of AEF.

Module	Technical Points	Problems Solved	Technical Approach	Potential Limitations and Applicability Boundaries
Data preprocessing	Unified resampling and normalization of multi-source Data	Multi-modal data show significant differences in resolution and units	Unify multi-source raster data to a common spatial scale, normalize separately by data source; retain sensor type information as conditional input	Details from high-resolution data sources may be lost, and there is a risk of averaging out the intensity of different sensor information
	Spatiotemporal alignment and observation window reorganization	Multi-source data have irregular temporal observations, with prominent missing data across sources	Construct variable-length time series, distinguishing between input time range and output aggregation time range	Limited sensitivity to short-term sudden events, weaker stability in areas with long-term missing data
	Quality control and noise processing	Interferences such as clouds, shadows, and speckle noise hinder effective learning	Combine quality masks and geometric parameters to filter out low-quality reconstruction targets	Quality screening rules are complex and not fully disclosed, potentially introducing bias in available data
	Global spatial balanced sampling	Samples are concentrated in densely populated human activity areas, while some ecological regions	Stratified sampling based on ecoregions, with constraints on sample spatial proximity	Weaken the depiction of local heterogeneous areas and rare ecological types

		have scarce samples		
	Aligning geographic text with species data	Pure remote sensing data struggle to capture ecological and semantic hierarchical information	Mapping Wikipedia entries and GBIF records into a unified spatial grid for training	Affected by human observation bias, there is only weak alignment between semantics and pixels
Model training	Cross-source reconstruction loss	In scenarios without annotations, models tend to learn “cross-source shortcuts”, leading to representation degradation	Using cross-source reconstruction as the main task, leveraging partial modalities and temporal conditions to recover other observations	Reconstruction performance does not fully align with downstream task performance, with limited generalization to new modalities
	Batch uniformity constraint	Imbalanced data distribution causes models to favor high-frequency regions and common patterns	Introducing uniformity constraints in training batches to stabilize embeddings in spherical space	May weaken the expression of real-world spatial imbalances
	Student-teacher consistency training	Training is easily affected by temporal fluctuations and data perturbations, leading to representation drift	Using Exponential Moving Average (EMA) teacher to enforce embedding consistency across different views	Teacher bias may accumulate, resulting in slower adaptation to new distributions
	Remote sensing-text contrastive learning	The semantic hierarchy of remote sensing features is insufficient to associate ecological and species concepts	Semantic weak anchoring is achieved through contrastive learning between remote sensing embeddings and text embeddings	Text noise may interfere with training, and the interpretability of semantic associations is limited
	Spatial-Temporal Path (STP) encoder	It is necessary to consider spatial structure,	A parallel encoding approach using spatial, temporal, and accuracy	The architecture complexity and parameter scale are

		temporal dynamics, and local accuracy simultaneously; a single path is difficult to handle this comprehensively	paths with mid-level fusion is adopted	relatively large, and its advantages are limited under extremely sparse temporal conditions
Representation formation	Von Mises-Fisher (VMF) spherical bottleneck compression	High-dimensional embedding storage and computational costs are high, and distribution stability is insufficient	Use VMF distribution constraints and compress representations into a low-dimensional spherical space	Compression may lose some fine-grained information, which is unfavorable for fine-grained inversion tasks
	Unified spherical latent space	Embedding comparability between regions and tasks is weak, making cross-scenario generalization difficult	Map all pixels to a unified spherical latent space	Spherical distances may not strictly correspond to real physical differences
	Time-conditioned annual embedding output	Downstream tasks have different temporal scale requirements, and it is necessary to flexibly generate corresponding representations	Generate annual or multi-annual embeddings based on a given time window	Long-term aggregation may smooth out short-term anomalies, and extrapolation capability is limited
	No fine-tuning required for downstream usage	Users lack labeled data and computational power, making complex fine-tuning difficult	Downstream tasks can directly use linear models, regressors, or KNN for adaptation	Complex nonlinear tasks still require dedicated modeling

4. Literature Review

4.1. Survey Scope

As of Feburary 29, 2026, this study conducted a systematic search using the keyword "AlphaEarth" on Google Scholar, the official AlphaEarth platform, Web of Science, and ResearchGate, with the search scope limited to title, abstract, and keywords fields. After manual screening, news reports, blog posts, technical interpretations, and other non-research texts were excluded, resulting in the inclusion of 23 research articles, including the original paper published by the DeepMind team. The search results indicated that, as of this date, no systematic review or methodological synthesis on AEF has been found, which directly justifies the present review. The list of included literature is provided in Appendix A.

In terms of overall characteristics, existing AEF research remains in an early diffusion phase: on one hand, knowledge production follows a pattern where it is initially proposed by a few industry teams and subsequently rapidly adopted by universities and public research institutions; on the other hand, preprints dominate as the publication medium, indicating that this field is currently primarily characterized by rapid prototyping, method validation, and scenario exploration. As of the search date, no systematic review or methodological integration on AEF has been identified. institutional types The composition of institutions, distribution of publication venues, and the distribution by country of the relevant literature are shown in Figure 2.

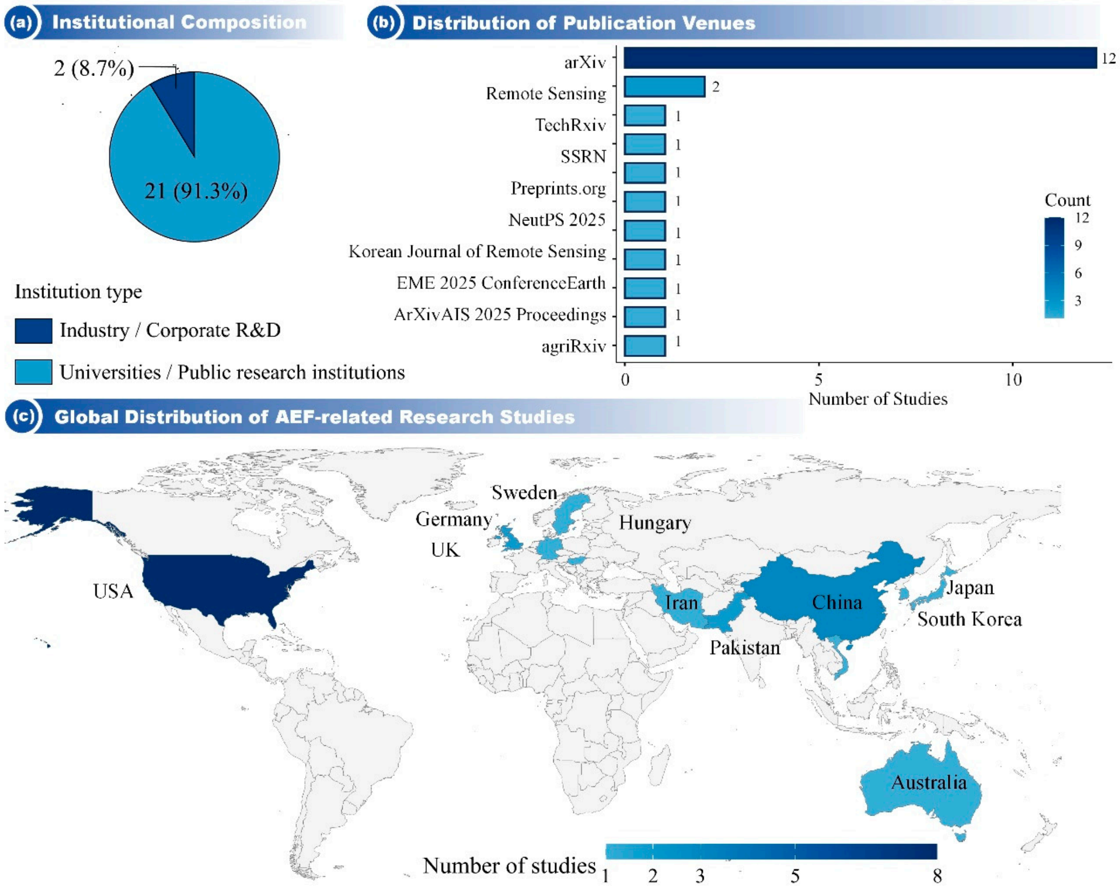

Figure 2. Statistics on publication venues, country distribution, and institutional types of AEF-related research studies.

4.2. Research Topics & Regions

As AEF is a global-scale EO fundamental embedded product with extensive spatial coverage and potential application scenarios, it is necessary to systematically review the application topics and regional distributions of existing studies to identify spatial gaps, thematic biases, and their potential causes in current research.

4.2.1. Research Topics

Existing studies have covered multiple fields, including agriculture, ecology, hydrology, disaster management, urban environments, and socioeconomics. In agriculture, research focuses on the classification of cool-climate crops such as corn, wheat, and soybeans (Murakami, 2025), identification of land cover types like wetlands and croplands (Böröcz & Molnár, 2025; Khan & Ahmad, 2025; Ryan et al., 2026; Zvonkov et al., 2025), mapping of peanut and grain intercropping systems (Lisaius et al., 2026), benchmarking agricultural tasks (Ma et al., 2026), and crop yield estimation (Fang et al., 2025). In ecological and forest studies, topics include vegetation mapping (Houriez et al., 2025), estimation of above-ground forest biomass (Lucero et al., 2026), and temporal mapping of canopy height (Ngo et al.). In water environment and hydrology, research covers estimation of river water quality indicators (Kim et al., 2025), runoff prediction (Qu et al., 2026), and flood inundation forecasting (Ashfaq et al.). Disaster-related studies involve mapping and change detection of burned areas from wildfires (Seydi, 2025), analysis of building damage in wildland-urban interface zones (Esparza et al., 2025), and landslide susceptibility mapping (Cheng et al., 2026). Urban and socio-environmental aspects include prediction of air pollutants (Alvarez et al., 2025; Fang et al., 2026), spatial prediction of health indicators in low- and middle-income countries (Metz et al., 2025), poverty index mapping (Pettersson & Daoud, 2025), and urban semantic representation (Liu et al., 2025).

Overall, existing studies are clearly concentrated on categorical and structural topics, such as crop classification and land cover mapping. These tasks align well with the annual static expression of AEFs, which are inherently adept at capturing long-term stable features such as vegetation structure and surface cover. Classification and zoning tasks typically require lower temporal precision, making them the most numerous and earliest applications. In contrast, continuous variable and process-driven tasks are relatively fewer, such as crop yield estimation, biomass assessment, runoff prediction, and air pollution forecasting. These tasks not only depend on surface spatial structures but are also significantly influenced by meteorological, hydrological, and biophysical processes, requiring higher temporal resolution and sensitivity to external driving factors. Since annual AEF embeddings primarily represent long-term background characteristics, they often fail to provide a natural advantage in these tasks and are better suited as supplementary representations within multi-source information fusion frameworks. Socioeconomic and public health topics are even more limited, as their relationships with surface structures are mostly indirect and usually require integration with socioeconomic statistics or other semantic data, thereby increasing research complexity and entry barriers.

4.2.2. Research Regions

Although existing studies have covered the Americas, Europe, Asia, Africa, and Oceania, their spatial distribution remains highly uneven, with a clear focus on the United States and its related regions, while other areas are mostly represented by single-country or localized case studies. The most concentrated research is in North America, including the contiguous United States (Ma et al., 2026), Michigan (Murakami, 2025), MTBS (Seydi, 2025), 13 states in the Corn Belt (Fang et al., 2025), CAMELS-US watersheds (Qu et al., 2026), the Eaton Fire area in Los Angeles County, California (Esparza et al., 2025), and cross-border studies that use northern U.S. and Alaska as training domains and extrapolate to southern Canada (Houriez et al., 2025). In Europe, studies include Germany (Murakami, 2025), Emilia-Romagna in Italy (Cheng et al., 2026), European fire monitoring sites (Seydi, 2025), Hungary (Böröcz & Molnár, 2025), and London(Liu et al., 2025). Asian studies are distributed across Hokkaido, Japan (Murakami, 2025), the four major river systems of South Korea (Kim et al., 2025), Changchun, China (Fang et al., 2026), Hong Kong, Nantou County, Taiwan (Cheng et al., 2026), Cuc Phuong National Park in Vietnam (Ngo et al.), and Pakistan (Ashfaq et al.; Khan & Ahmad, 2025). African studies cover the peanut basin in Senegal (Lisaius et al., 2026), Malawi (Metz et al., 2025), Togo (Zvonkov et al., 2025), and 25 countries in sub-Saharan Africa (Pettersson & Daoud, 2025). Latin American studies include the Las Piedras watershed in Cauca Province, Colombia (Lucero et al., 2026), and the Quito Metropolitan Area in Ecuador (Alvarez et al., 2025), while only the Narran Lake wetlands in Australia (Ryan et al., 2026) are included in Oceania.

Further analysis reveals a strong problem-oriented relationship between regional distribution and research themes, as shown in Figure 3. The concentration of studies in the United States is closely associated with its relatively well-developed remote sensing, statistical, and ground observation data systems, making it frequently used for tasks such as agricultural mapping, watershed runoff prediction, and disaster risk assessment. Cool-climate agricultural regions like Germany and Hokkaido have become typical scenarios for trans-regional crop classification studies; the four major river systems in South Korea face long-term challenges from eutrophication and water environment management, thus creating clear application needs for water quality index estimation. Changchun, a northeastern Chinese city affected by winter heating, industrial emissions, and meteorological conditions, has strong practical significance for ultra-high-resolution air pollution exposure mapping. The peanut basin in Senegal, characterized by fragmented agricultural plots and high landscape heterogeneity, serves as a typical region for testing the applicability of embedded representations in complex smallholder landscapes. Studies in Malawi and 25 countries in sub-Saharan Africa focus on health service coverage and poverty indices, reflecting greater attention to public health and socio-development issues in these regions. Pakistan, prone to flash floods and featuring complex land use patterns, has become an important context for hydrological disaster prediction and land cover studies. The Andean region in Colombia and the national parks in Vietnam, due to their high biodiversity and complex forest structures, are key areas for estimating above-ground forest biomass and tree height. Although AEF-related research has shown a trend toward trans-regional expansion, its spatial distribution still clearly favors regions with better data foundations and more mature validation conditions. By contrast, research in tropical regions, arid zones, high-latitude cold regions, and the Southern Hemisphere remains relatively insufficient.

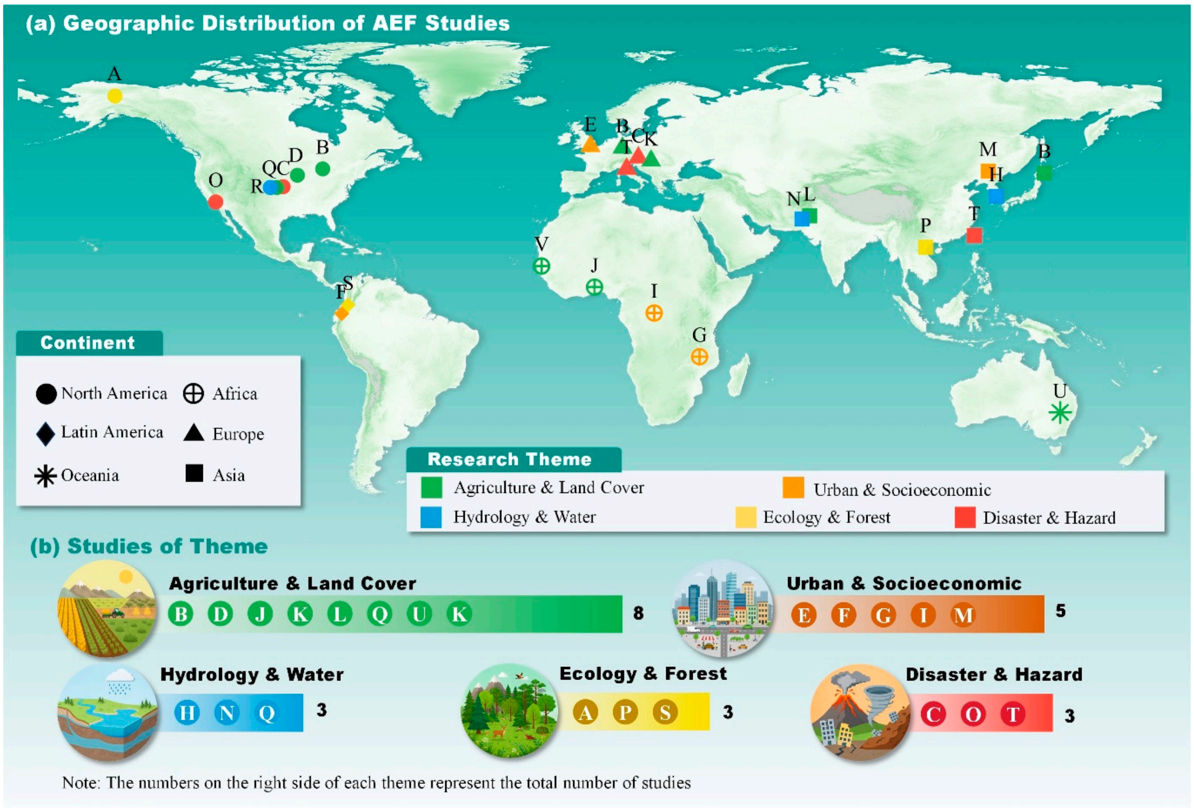

Figure 3. Global geographic distribution and thematic classification statistics of AEF research (letters in the figure correspond to the Letter ID of papers in Appendix A).

4.3. Core Scientific Questions

The 22 empirical studies included, although covering diverse fields such as agriculture, ecology, hydrology, disaster, urban and socio-economic aspects, show a high degree of consistency in the scientific questions they address. These questions all revolve around the generalization capability, information boundary, and application value of AEF as an annual-scale static surface embedding. Overall, these studies have generally gone through three stages: early work primarily addressed the question of “whether it can be used directly,” i.e., whether AEF could support regional-scale classification and estimation without complex preprocessing, feature engineering, or phenological alignment; subsequent research shifted toward “generalization mechanism identification,” focusing on stability and failure conditions in cross-regional, cross-year, and cross-scale applications; recent studies have increasingly explored “structured enhancement,” attempting to improve the interpretability, reliability, and deployment capability of AEF by introducing semantic information, spatial relationships, physical constraints, or multi-source fusion. In summary, AEF research has gradually transitioned from initially focusing on “whether it can improve indicators” to addressing “under what conditions it is effective and how its capability boundaries can be expanded.”

Based on this, existing research can be summarized into six interrelated core questions: (1) Can AEF directly replace traditional workflows? (2) Under what transfer conditions does it fail? (3) What are the information capacity and dimensional structure within the embedding? (4) Does the annual-scale static embedding inherently have limitations when dealing with tasks that depend on short-term changes or dynamic processes? (5) When a single embedding is insufficient, how should enhancements be achieved through semantics, spatial relationships, or physical constraints? (6) Does AEF possess compensatory capabilities under conditions of data scarcity, sparse observations, or incomplete labels?

4.3.1. Plug-and-Play Usability

The first main thread is whether AEF can be used directly. The study treats AEF as a pre-generated general embedding layer and, with minimal introduction of complex preprocessing, spectral index construction, or phenological alignment, directly combines it with tree models, linear models, or other lightweight classifiers to complete classification and estimation tasks. This aims to test whether AEF can serve as a “plug-and-play” feature foundation by shifting the traditional remote sensing workflow forward. Such explorations are most typical in agricultural and land cover mapping tasks because AEF itself is trained on multi-temporal natural surface observations, and its embeddings inherently encode information about vegetation structure, surface types, and phenological differences. Therefore, it is prioritized for classification and estimation problems related to natural surface properties.

For example, studies have systematically compared the performance of simple models combined with AEF in regional vegetation and crop classification, discussing whether model accuracy remains stable after reducing manual feature design (Houriez et al., 2025); some research has directly tested whether an “embedding + lightweight classifier” approach can support rapid large-scale mapping iterations and discussed how to handle probabilistic outputs with high recall but excessive false positives (Zvonkov et al., 2025); under conditions of limited samples or scarce annotations, other studies have explored the feasibility of object recognition based solely on embedding similarity or semantic prototypes (Böröcz & Molnár, 2025). At the same time, some research further places AEF embeddings alongside specifically designed remote sensing feature engineering models within the same framework for comparison, examining whether stable accuracy can be maintained while reducing deployment complexity (Ma et al., 2026); and starting from traditional remote sensing workflows, others have questioned whether AEF can achieve the accuracy of existing processes without cumbersome preprocessing and whether it brings more coherent spatial structures with smoother boundaries (Ryan et al., 2026).

4.3.2. Transfer Boundaries

The second major research thread concerns the question of “Under what conditions does AEF transfer fail?”. While the issue of direct usability examines whether a model can be applied directly within the same region and setting, this line of inquiry further explores how performance changes when tasks cross regions, years, or spatial scales, as well as which factors trigger degradation in

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 May 2026 doi:10.20944/preprints202605.0981.v1

performance. These investigations primarily focus on three aspects: cross-regional extrapolation, cross-year application, and scale transfer, aiming to identify sources of instability in transfer experiments.

In terms of cross-regional extrapolation, studies mainly examine the generalization boundaries of models when transferred from one ecoregion, country, or continent to another. For example, North American vegetation mapping research extended a model trained in northern U.S. and Alaska to southern Canada to test migration stability across different vegetation classification levels (Houriez et al., 2025); a cool climate crop classification study conducted cross-testing among Germany, Hokkaido (Japan), and Michigan (U.S.), examining whether the same crop types exhibit consistent transferability across countries (Murakami, 2025); wildfire detection research further extrapolated a model trained on U.S. MTBS data to fire sites in Europe to evaluate its generalization capability under intercontinental conditions (Seydi, 2025); landslide susceptibility mapping (Cheng et al., 2026) and catchment runoff prediction (Qu et al., 2026) studies tested model performance between different regions and unobserved watersheds, respectively, to assess the impact of geographic context differences on transfer stability. In terms of cross-year application, the focus shifts to the influence of temporal mismatch on model stability. A typical example is the smallholder farming area crop mapping study in Senegal, which conducted bidirectional experiments between 2018 and 2019 and forward extrapolation to 2021 to examine the model's generalization ability for the same crop type across different years (Lisaius et al., 2026). In terms of scale transfer, studies primarily investigate whether models remain effective across different spatial units. Agricultural mapping benchmark research conducted spatial transfer experiments across different states and ecoregions in the U.S., while comparing transfer performance between county-level and field-level scales, discussing whether models trained at coarser statistical scales can generalize to finer plot-level scales (Ma et al., 2026).

4.3.3. Information Structure

The third main research thread focuses on how much effective information AEF encodes and how this information is distributed in the embedding space. Unlike the first two threads, which directly compare task accuracy, this line of research emphasizes evaluating the information capacity, dimensional redundancy, and structural organization of embeddings through reference frameworks. Studies actively construct a "reference framework" for comparison, such as index systems based on domain knowledge, sets of physical variables, LiDAR-measured tree heights, or long-term water quality monitoring data, serving as standard feature sets containing key mechanistic information. These specialized variables are first used to train traditional models to establish performance benchmarks, then compared with models using only AEF embeddings. If AEF achieves comparable results without introducing these physical or mechanistic variables, it indicates that the embeddings implicitly encode these critical structural features; if significantly inferior, it suggests that certain biophysical details or process variables remain underrepresented (Lucero et al., 2026). In tree height mapping, studies further use limited ground plots or GEDI LiDAR sample points as external references and compare whether annual results maintain structural consistency to assess whether the embeddings stably represent forest vertical structure (Ngo et al.).

In addition to information capacity, these studies also focus on the internal organization of information within embeddings. For example, in multi-index water quality estimation, they analyze whether important dimensions associated with different indices overlap to determine whether a relatively stable "aquatic information subset" exists; in landslide susceptibility mapping, they compare the performance differences between full 64-dimensional embeddings and principal component analysis (PCA)-reduced dimensions to test whether information is concentrated in a few principal components and whether significant redundancy exists (Cheng et al., 2026; Kim et al., 2025). Additionally, some studies further assess whether embeddings miss regional information by examining residual spatial distributions, such as analyzing whether prediction residuals exhibit spatial clustering in crop yield forecasting to identify whether AEF fails to adequately encode controlling factors like management practices or climatic gradients (Fang et al., 2025). Overall, this research thread is no longer solely concerned with predictive performance but instead decomposes the information capacity, organizational patterns, and potential boundaries of AEF through reference frameworks, dimensional structures, and residual diagnostics.

4.3.4. Temporal Sensitivity

The fourth main thread discusses whether annual static embedding are inherently limited when tasks clearly depend on short-term variations or dynamic processes. AEF, by nature, is a compressed representation of long-term surface structures and is more suitable for depicting stable spatial backgrounds; however, many practical tasks are simultaneously influenced by short-term weather conditions, sudden disturbances, or rapid processes. Therefore, this line of inquiry no longer merely compares generalization performance across different regions but further distinguishes between "structural factors" and "processual factors," examining whether annual embeddings remain effective as the dynamic components in tasks increase.

For example, in event change detection, the study directly tests whether annual summaries of wildfire burn scars and change detection dilute event-level change signals by comparing two strategies: "explicit differencing" and "implicit learning." It also examines whether differencing operations can re-activate compressed change information, thereby defining the expressive boundaries of AEF in identifying sudden disturbances (Seydi, 2025). Furthermore, in air pollution research, the focus shifts from merely identifying changes to determining which types of driving mechanisms AEF is better suited to express. By comparing different pollutants side by side, the study indicates that AEF is more sensitive to long-term structural variables such as road density and industrial layouts, while its performance is limited for pollutants that rely more heavily on short-term meteorological and chemical reaction processes (Alvarez et al., 2025). Building on this, $PM_{2.5}$ exposure mapping advances the issue to a finer spatial scale, focusing on whether annual embeddings can still capture neighborhood-level concentration gradients in the absence of high-frequency covariates such as meteorological variables and Aerosol Optical Depth (AOD), and discussing the trade-offs among accuracy, bias, and spatial stability (Fang et al., 2026). Landslide susceptibility mapping further incorporates "temporal consistency" into the experimental design, comparing prediction results from embeddings of different years to examine whether AEF primarily expresses long-term surface structures or can maintain some sensitivity to event-triggering conditions specific to certain years (Cheng et al., 2026).

4.3.5. Information Supplementation

The fifth main thread focuses on how to strengthen a single AEF embedding when it is insufficient to support complex reasoning by introducing external information. An AEF is essentially a compressed representation of natural surface observations, better suited for characterizing vegetation structure, land cover types, and long-term background differences. However, higher-level information such as spatial dependencies, functional attributes, and process constraints cannot be explicitly expressed within a single pixel vector. For this reason, the research no longer treats AEF merely as an independent feature but instead attempts to place it within a more comprehensive reasoning framework. By supplementing semantic information, spatial relationships, and mechanistic constraints, the limitations in its expression during complex tasks can be addressed.

Such supplementation mainly manifests in three aspects: first, introducing external semantic information, such as Point of Interest (POI) data in urban studies, to enhance differentiation of functional attributes (Liu et al., 2025); second, explicitly modeling spatial relationships, such as incorporating graph structures in poverty or settlement studies to express spatial dependencies between adjacent units (Pettersson & Daoud, 2025); third, adding physical constraints or propagation structures to ensure that model outputs conform to basic process rules (Ashfaq et al.) and distinguish between background risks and neighborhood effects (Esparza et al., 2025). Overall, this main thread is not solely concerned with improving accuracy but rather with addressing the expressive limitations of AEF in complex reasoning scenarios by supplementing semantic, relational, and mechanistic information.

4.3.6. Compensatory Capability

The sixth main theme focuses on the question: “Can AEF serve as a compensatory mechanism when the original data system has gaps.” In real-world applications, research rarely operates under ideal conditions with complete variables, sufficient labels, and dense observations; instead, it often faces sparse observations, missing variables, label noise, and scale mismatches. Therefore, this theme does not examine whether AEF is optimal, but rather investigates whether it can provide a relatively stable contextual representation to compensate for structural gaps in the data system when key data are insufficient.

In current research, the compensatory role of AEF is primarily evident in three scenarios: First, under sparse observation conditions, it provides additional spatial context for tasks such as health and environment. For example, in health-related tasks in low- and middle-income countries, where monitoring stations are scarce, health data coverage is limited, and traditional spatial interpolation methods face constraints, studies have tested whether embedding can serve as a supplementary foundation to support more reliable spatial estimation (Metz et al., 2025); Second, even in the absence of high-frequency covariates, it helps models capture long-term spatial structural differences. For instance, in $PM_{2.5}$ exposure estimation, studies specifically test AEF-only models to assess whether embeddings can maintain relatively stable predictive performance and control bias without incorporating high-frequency covariates such as meteorological variables and AOD (Fang et al., 2026); Third, when labels are incomplete, probabilistic outputs are uncertain, or spatial structures are fragmented, it improves the stability and spatial coherence of results. For example, national-scale cropland mapping studies discuss how threshold strategies or probability adjustments can be used to make outputs more reasonable in terms of area estimation and spatial distribution when there is high recall but excessive false positives (Zvonkov et al., 2025), while high-resolution mapping studies further alleviate “salt-and-pepper noise” and boundary fragmentation issues at the pixel level through object-level aggregation or structural constraints (Khan & Ahmad, 2025; Kim et al., 2025). The value of AEF in these tasks does not necessarily lie in achieving the highest accuracy, but rather in maintaining the usability, stability, and controllability of bias under data-limited conditions.

4.4. AEF Usage Patterns

After clarifying the core scientific questions, it is even more critical to understand the specific usage patterns of AEF, including its role, scale, and model compatibility; the detailed content framework is shown in Figure 4. This systematic analysis at the “usage level” helps reveal the strengths and limitations of AEF and provides a basis for its expansion across different tasks.

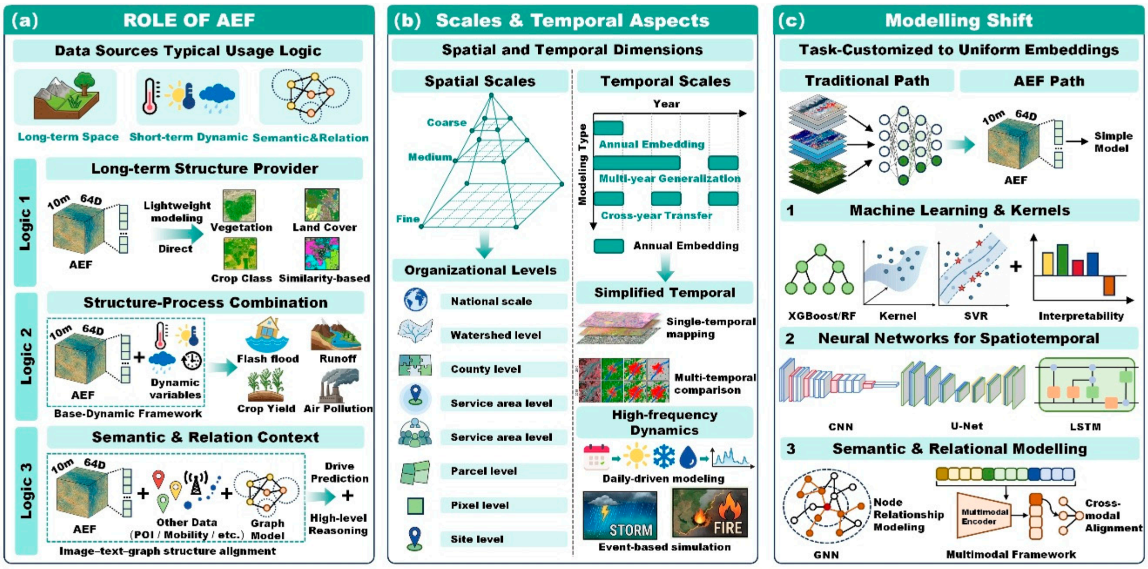

Figure 4. Comprehensive Framework of AEF Usage Patterns.

4.4.1. Roles

From the perspective of information sources, remote sensing tasks typically rely on three types of information. The first is long-term spatial structural information, used to characterize relatively stable natural differences on the Earth's surface, such as topography, vegetation, and land cover, explaining "why some places are inherently different." The second type is short-term or dynamic driving information, such as rainfall, temperature, and extreme events, explaining "why changes occur at a certain time." The third type is semantic and relational information, such as spatial adjacency and propagation mechanisms, used to answer questions like "who influences whom" in structured reasoning. As an annual-scale surface embedding, AEF's training mechanism determines its fundamental proximity to the compressed expression of long-term spatial structural information. However, during the current exploratory phase of applications, its specific functional positioning remains inconsistent. Integrating existing research, we can roughly categorize it into three typical usage logics.

The first category uses AEF as a provider of long-term spatial structural information. This approach is the most direct and logically the "lightest." AEF, as a 64-dimensional annual embedding, is used as the sole or primary input feature without introducing additional explicit variables, directly completing classification or regression tasks. Examples include vegetation type mapping (Houriez et al., 2025), multi-regional crop classification (Murakami, 2025), and national-scale cropland identification (Ma et al., 2026; Zvonkov et al., 2025), all achieving good results. It can also replace traditional indices and phenological features for wetland (Ryan et al., 2026) or land cover mapping (Lisaius et al., 2026), even enabling class discrimination through embedding similarity (Böröcz & Molnár, 2025). These tasks are usually dominated by long-term spatial patterns, with AEF supporting predictions and significantly reducing feature engineering complexity. However, this method has limitations in tasks strongly influenced by short-term drivers. For instance, in air pollution studies, it performs well for pollutants related to road density and industrial layouts but shows weaker explanatory power for indicators more dependent on meteorological and chemical processes (Alvarez et al., 2025).

The second category combines AEF as an expression of both long-term structural information and dynamic driving information. When the study subject is simultaneously affected by spatial structure and temporal change, AEF is often used together with meteorological, hydrological, or human activity data to form a "structure-process" framework: AEF provides spatial differences, while dynamic variables explain temporal changes. For example, in debris flow prediction, it serves as basic underlying surface information combined with rainfall and soil moisture to distinguish flood sensitivity across regions (Ashfaq et al.); in catchment runoff simulation, it acts as a compressed expression of static catchment attributes combined with daily meteorological data to explain response differences among catchments (Qu et al., 2026); in air pollution modeling, it represents the long-term environmental context within cities, combined with AOD and meteorological variables to differentiate spatial variations and temporal fluctuations (Fang et al., 2026); in crop yield estimation, AEF captures regional long-term differences, while interannual variation is mainly driven by climatic conditions (Fang et al., 2025). In this category, AEF no longer explains all variability alone but functions as a "spatial structure base," collaborating with dynamic variables in a division of labor.

The third category incorporates AEF as part of the environmental component within a semantic and relational information framework. In tasks involving urban function identification, poverty mapping, public health, or disaster spread, spatial structure and short-term driving information alone are insufficient for reasoning; thus, AEF is embedded into more complex semantic or graph-based frameworks, participating in reasoning as part of the environmental information. For example, it combines with POI text "functional labels" as "image structure" to define urban functional zones (Liu et al., 2025); in health service prediction (Metz et al., 2025), poverty mapping (Pettersson & Daoud, 2025), and wildfire damage prediction (Esparza et al., 2025), it integrates with digital behavior, mobility data, and building structure attributes, respectively, serving as node attributes in graph models to enable reasoning. In these tasks, AEF provides background semantics at the environmental level, but the actual prediction outcomes are determined by semantic alignment or relationship modeling mechanisms. Compared to the previous two categories, here AEF no longer merely explains “where differences exist” or “why changes occur,” but instead provides environmental context for higher-level structured reasoning.

4.4.2. Scales

From the spatial scale perspective, the research simultaneously manifests in two dimensions: resolution and spatial organizational hierarchy. In terms of resolution, most studies directly adopt the native 10m resolution of AEF for pixel-level mapping, emphasizing wall-to-wall coverage and precise boundary delineation; some extend to coarser scales, such as 30m resolution (Cheng et al., 2026), county units (Fang et al., 2025; Ma et al., 2026), or watershed scales (Qu et al., 2026), achieving regional-level analysis and generalization through feature aggregation; others explore extensions toward object-level units, such as building-scale (Esparza et al., 2025) or small parcel identification (Lisaius et al., 2026). In terms of spatial organizational hierarchy, tasks span from pixel/parcel level to community or point-level, county-level, service areas, watersheds, and even national scales. Overall, AEF can be directly used for pixel prediction at fine resolutions or adapted to different decision-making units through spatial aggregation, demonstrating strong scale adaptability.

From the temporal scale perspective, most studies use annual embedding as the basic time unit, emphasizing cross-year migration or multi-year generalization capabilities; some tasks employ single-period mapping (Lucero et al., 2026; Ryan et al., 2026) or dual-time comparisons (Seydi, 2025), simplifying the temporal dimension; a few studies introduce higher-frequency dynamic inputs, such as daily meteorological-driven runoff modeling (Qu et al., 2026) or event-level flood and fire simulations (Ashfaq et al.; Esparza et al., 2025). Most works do not model continuous temporal processes but instead treat AEF as a structural background at the annual scale. Overall, temporal modeling primarily focuses on “annual structure representation + cross-year generalization,” while high-frequency dynamic process modeling remains concentrated in a few tasks.

4.4.3. Models

Before the emergence of AEF, typical remote sensing tasks usually relied on increasingly complex and highly customized technical pathways. These included extensive manual feature engineering (e.g., multi-temporal indices, phenological parameters, harmonic decomposition, texture, topographic factors), integration of multi-source and multi-modal data (e.g., optical, SAR, LiDAR, meteorological, statistical data), and deep network architectures (e.g., temporal CNNs, ConvLSTMs, multi-stream networks, self-attention modules) that were continuously stacked. Models typically required custom-designed input systems and structures tailored to specific regions and variables, resulting in high engineering costs and limited transferability. The proposal of AEF itself aimed at “expression prepositioning and process simplification.” It compresses multi-year, multi-source, and multi-modal data into a unified 10m, 64-dimensional annual embedding, completing temporal alignment and feature extraction in advance, thereby significantly reducing the reliance of downstream tasks on complex feature engineering and specialized models. Under this premise, current research shows a clear shift in technical orientation: traditional machine learning models are predominant, with deep models as supplements; model complexity shifts from “structural involvement” to “information division.” This pattern marks a transition from “customizing complex models for each task” to “completing tasks with as simple models as possible on a unified embedding,” achieving convergence and simplification at the engineering level.

Firstly, the largest proportion consists of traditional machine learning models such as tree models and kernel methods. Random Forest (RF), Extreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), and Support Vector Regression (SVR) appear as primary models in most tasks. These models are insensitive to feature scales, capable of capturing certain nonlinear relationships, have low training costs, and offer good interpretability when combined with methods like SHapley Additive exPlanations (SHAP). In scenarios using only AEF, its embedding can be directly used as input for classification or regression; when combined with dynamic variables, tree models also serve as the backbone framework, distinguishing between "spatial structural differences" and "temporal change factors" through feature importance or residual analysis. Compared to the previous approach relying on complex feature engineering or deep networks before AEF's emergence, this paradigm significantly reduces modeling complexity, bringing traditional models back to the mainstream.

Secondly, in tasks involving explicit temporal evolution or spatial dependence, corresponding neural network structures are generally introduced. Tree models essentially perform independent mapping per sample and cannot model neighborhood interactions or temporal memory, making them unsuitable for continuous fields or dynamic processes. For example, in detecting burned areas from wildfires (Seydi, 2025), changes often occur in contiguous patches with clear boundaries. When tree models predict pixel by pixel, spatial consistency requires post-processing correction; whereas U-Net takes an entire region as input and naturally maintains spatial coherence in its output, directly generating continuous probability maps. In flood prediction (Ashfaq et al.), water spreads along low-lying terrain areas, which is essentially a neighborhood propagation process. Tree models can only indirectly express this through constructed flow accumulation indices, while convolutional networks continuously integrate neighborhood information during forward propagation, enabling diffusion relationships to form internally within the model; simultaneously, end-to-end differentiable structures allow physical constraints to be incorporated into the loss function, achieving joint optimization of physical consistency and predictive accuracy. In catchment-scale runoff prediction (Qu et al., 2026), the key lies in temporal accumulation effects. Tree models treat daily samples independently and cannot remember prior rainfall or soil moisture states unless lagged variables are explicitly constructed; in contrast, LSTMs retain historical information via internal state mechanisms and can automatically learn responses across days or even weeks.

In scenarios involving semantic relationships or complex structure modeling, models expand into graph neural networks or multi-modal frameworks, where the problem is no longer point-wise prediction but rather node relationship modeling or cross-modal alignment. In poverty prediction (Pettersson & Daoud, 2025), graph convolutional networks use AEF as node attribute inputs and introduce spatial adjacency matrices to aggregate neighbor node information and learn inter-regional influences. Predictions are no longer based on isolated samples but propagate and update over the graph structure. Wildfire damage prediction first employs graph networks to learn complex structural interaction representations, then feeds the outputs as high-level features into XGBoost, allowing the tree model to complete the final prediction by combining "structural expression capability" with "nonlinear mapping capability." In urban function identification, the focus shifts to cross-modal alignment, where a lightweight contrastive learning framework aligns AEF embeddings with POI text embeddings in vector space, with predictions based on the aligned joint representation. A common characteristic of these models is that AEF is no longer a single input feature but is embedded into higher-order relational structures or semantic spaces, participating in node propagation, attention allocation, or cross-modal alignment. Predictions depend on the integrated expression of "relationships–structures–semantics."

4.5. Performance Metrics

The performance evaluation systems adopted in existing studies for applications involving AEF embedding can generally be categorized into four types: first, task performance metrics, used to assess classification and regression outcomes; second, distribution and spatial structure consistency metrics, used to examine the rationality of map outputs in terms of statistical distribution, spatial patterns, and temporal trends; third, error and robustness diagnostic metrics, used to analyze residual structures, feature contributions, and cross-regional generalization fluctuations; fourth, physical and quality control metrics, used to constrain the physical plausibility and cartographic credibility of results. The relevant metrics and their application scenarios are summarized in the Table 3 below.

Table 3. Summary of performance metrics and their application scenarios.

Dimension	Type	Index	Definition	Paper Application
Task Performance	Classification Accuracy	Accuracy、Overall Accuracy (OA)	The proportion of samples correctly classified by the model	Measures the overall discriminative ability for vegetation and land cover (Böröcz & Molnár, 2025; Houriez et al., 2025; Khan & Ahmad, 2025; Ryan et al., 2026; Zvonkov et al., 2025), crop classification (Lisaius et al., 2026; Murakami, 2025), damage segmentation detection (Esparza et al., 2025; Seydi, 2025), and landslide susceptibility mapping (Cheng et al., 2026)
		User's Accuracy (UA) /Precision	The proportion of samples predicted as a certain class that truly belong to that class, reflecting false positives	Decomposes missed detections and false alarms in per-crop classification (Murakami, 2025), measures recall level and missed judgments in cropland mapping, and analyzes the tendency of high recall accompanied by increased false alarms (Zvonkov et al., 2025)
		Producer's Accuracy (PA) /Recall	The proportion of actual samples of a certain class that are correctly predicted, reflecting missed detections	
		F1	The harmonic mean of precision and recall, comprehensively measuring the performance of minority class identification, suitable for imbalanced class situations	Measures the balance between precision and recall for minority pixels in burn area segmentation (Seydi, 2025), cropland mapping (Zvonkov et al., 2025), building damage identification (Esparza et al., 2025), landslide risk area discrimination (Cheng et al., 2026), flood or inundation mask classification (Ashfaq et al.), and crop classification in agricultural benchmark tasks (Ma et al., 2026)
		Macro-F1	Calculates the F1 score for each class and then takes the average, emphasizing the performance of small classes	Measures balanced classification quality in land cover (Khan & Ahmad, 2025) and crop classification (Lisaius et al., 2026), avoiding dominance by large classes
		Weighted-F1	Calculates the F1 score with weighted consideration of	Measures balanced classification quality for wetland vegetation

		class frequencies, balancing overall accuracy and small class performance, suitable for cases with significant differences in class frequencies	(Ryan et al., 2026) and crop classification (Lisaius et al., 2026)
	MCC	A correlation coefficient based on the confusion matrix, with strong robustness	Complements OA and F1, measuring the comprehensive robustness of wetland vegetation classification under imbalanced conditions (Ryan et al., 2026)
Classification consistency	Kappa	Measure classification consistency, excluding random consistency	Use Kappa to measure the non-random consistency of crop classification (Murakami, 2025), the consistency of land cover classification (Khan & Ahmad, 2025), and the consistency of different workflows in wetland vegetation mapping (Ryan et al., 2026)
Classification ability	ROC-AUC	Assess model discrimination ability by the area under the curve; higher values indicate stronger discrimination ability, unaffected by thresholds	Evaluate the overall discrimination ability of landslide susceptibility risk scoring (Cheng et al., 2026)
	Specificity	Measures a model's ability to identify negative samples; higher values indicate fewer false positives	Measures the ability to identify non-landslide areas in landslide susceptibility mapping (Cheng et al., 2026)
Regression performance	R^2	Measures the explanatory power of regression models for data variance; the closer the value is to 1, the better the fit; values less than 0 indicate that the model performs worse than mean prediction	Measures the fitting and explanatory power of various regression tasks, including yield estimation (Fang et al., 2025), air pollution prediction (Alvarez et al., 2025), health facility output (Metz et al., 2025), river water quality estimation (Kim et al., 2025), poverty mapping (Pettersson & Daoud, 2025), $PM_{2.5}$ concentration mapping (Fang et al., 2026), canopy height mapping (Ngo et al.), agricultural baseline yield (Ma et

				al., 2026), and forest AGB inversion (Lucero et al., 2026)
		RMSE	Measures the difference between predicted and actual values, emphasizing large errors; smaller numerical values indicate more accurate models	Measures the magnitude of error in various predictive tasks, including yield estimation (Fang et al., 2025), air pollution prediction (Alvarez et al., 2025), river water quality estimation (Kim et al., 2025), $PM_{2.5}$ concentration mapping (Fang et al., 2026), tree height mapping (Ngo et al.), agricultural baseline yield (Ma et al., 2026), forest AGB inversion (Lucero et al., 2026), and landslide risk assessment (Cheng et al., 2026). Use RMSE to measure the magnitude of prediction error in yield estimation
		MAE	Measures the average absolute difference between predicted and actual values; MAE is insensitive to large errors and is suitable for evaluating overall error levels	Use MAE to measure typical errors in various predictive tasks, including air pollution (Alvarez et al., 2025), water quality estimation (Kim et al., 2025), poverty mapping (Pettersson & Daoud, 2025), AGB inversion (Lucero et al., 2026), and landslide risk assessment (Cheng et al., 2026)
Data distribution	Distribution consistency	L1 distance	Measures the average deviation between distributions or maps, calculating overall differences	Measures the overall distribution differences in urban functions or zoning results (Liu et al., 2025)
		Chebyshev distance	Measuring the maximum difference between distributions or maps, focusing on the worst deviation scenario	Measuring the maximum local deviation of urban functions or zoning results (Liu et al., 2025)
		KL divergence	Measuring information loss or deviation between distributions, indicating the	Measuring the degree of distribution information deviation for urban functions or zoning results (Liu et al., 2025)

			similarity of two distributions	
	Spatial object consistency	IoU	Measuring the overlap between predicted masks and ground truth, sensitive to shape and boundaries	Measuring the spatial overlap between predicted and actual areas in burned area segmentation (Seydi, 2025), and the spatial matching quality of flood inundation masks (Ashfaq et al.)
	Temporal structure and trends	Pixel-level slope	Characterizing inter-annual change consistency and rationality through the slope of trend lines	Measuring the consistency of inter-annual change trends in tree height mapping (Ngo et al.)
	Spatial structure and landscape pattern	Landscape pattern indices such as edge density, local entropy, and average patch area	Measuring the spatial structure, fragmentation, and continuity of maps, reflecting the spatial characteristics of the landscape	Measuring spatial structural and patch characteristic differences in wetland vegetation mapping (Ryan et al., 2026)
Error and efficiency diagnosis	Feature analysis	Moran's I	Examining the spatial clustering of model residuals, reflecting omitted spatial factors in the model	Measuring the spatial clustering degree of residuals in yield estimation (Fang et al., 2025)
		Feature importance (group)	Assessing the contribution of different information sources in the model	Assessing the relative contribution of different information sources in the building damage model (Esparza et al., 2025)
		Gain importance	In tree models, measuring the contribution of each feature to reducing prediction error	Assessing the error contribution of key embedding dimensions in water quality estimation (Kim et al., 2025)
		Permutation importance	Measuring the dependence of model predictions on features; the more important a feature, the more significant the performance decline	Assessing the model's dependence on key embedding directions in yield estimation (Fang et al., 2025)
	Robustness and uncertainty assessment	CV standard deviation, ±std	Measuring the performance fluctuations of the model across different datasets to reflect stability	Assessing the stability of health facility output prediction across different folds (Metz et al., 2025), and cross-national generalization fluctuations of MAE or R^2 in

				poverty mapping (Pettersson & Daoud, 2025)
	Process simulation and improvement	NSE	Measuring the degree of improvement of the model relative to mean prediction; higher NSE indicates better prediction	Measuring the degree of improvement of runoff prediction relative to the mean baseline (Qu et al., 2026)
		KGE	Comprehensively assessing the model's performance in terms of correlation, bias, and variation magnitude	Comprehensively assessing the fitting quality of runoff process simulation (Qu et al., 2026)
	Engineering usability evaluation	Computational efficiency, code complexity	Measuring model runtime efficiency and code complexity to reflect the ease of model deployment	Assessing the engineering usability level of land cover identification schemes (Böröcz & Molnár, 2025)
	Regression bias and probabilistic reliability	NMB、NME	Measuring systematic bias and overall error of the model	Measuring systematic bias and relative error levels of $PM_{2.5}$ predictions (Fang et al., 2026)
		Brier score	Measuring the match between probabilistic forecasts and actual outcomes	Measuring the calibration consistency of debris flow risk probability outputs (Ashfaq et al.)
Physical mechanisms	Physical mechanism consistency	Monotonicity compliance rate	Measuring whether the model meets monotonicity requirements; risk should not decrease when driver variables increase	Measuring whether the debris flow risk model satisfies physical monotonic constraints (Ashfaq et al.)
	Physical artifact suppression	False positive level in high HAND regions	Evaluating the effectiveness of artifact suppression in high topographic areas	Measuring the degree of topographic artifacts in debris flow risk maps (Ashfaq et al.)
	Qualitative consistency	Qualitative map consistency, spatial continuity inspection	Inspecting spatial continuity to ensure reasonable spatial distribution of features and avoid spatial fragmentation phenomena.	Measuring spatial coherence of vegetation or land cover mapping (Houriez et al., 2025), boundary rationality of feature identification results (Böröcz & Molnár, 2025), and spatial and trend rationality of crop mapping results (Lisaius et al., 2026)

5. Discussion

This paper discusses advantages, limitations, dimensional interpretability exploration, and future work prospects. The detailed framework is shown in Figure 5.

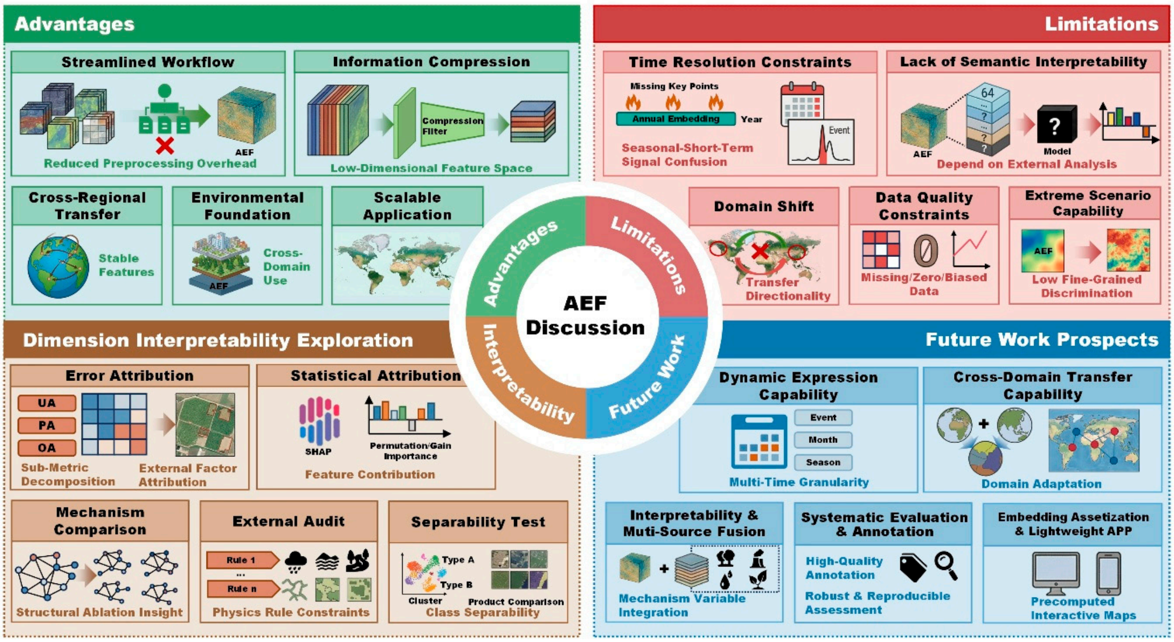

Figure 5. Comprehensive discussion framework of AEF's advantages, limitations, interpretability, and future directions.

5.1. Advantages

5.1.1. Streamlined Workflow

In large-scale mapping scenarios, AEF simplifies the workflow by fusing multi-source data and aligning temporal sequences into a unified embedded asset, thereby shifting certain data organization and preprocessing tasks that were previously repeated during the mapping phase to an earlier stage. Multiple studies have indicated that its embedded form allows downstream models to bypass the need for repeatedly constructing index systems or performing cumbersome preprocessing, enabling direct modeling to achieve stable results and significantly streamlining the process (Houriez et al., 2025; Metz et al., 2025; Murakami, 2025; Ngo et al.). In resource-constrained or rapid-experimentation contexts, this "pre-computed representation" approach demonstrates higher deployment efficiency and replicability (Lisaius et al., 2026). However, this advantage is primarily evident in general downstream tasks such as classification, identification, or mapping, rather than in fine-grained parameter inversion problems targeting specific mechanisms (Lucero et al., 2026).

5.1.2. Information Compression

AEF compresses multi-temporal, multi-modal observations into low-dimensional unified vectors, significantly reducing feature complexity while maintaining environmental semantic expressiveness. This low-dimensional embedding not only reduces computational burden but also enables traditional methods to achieve performance close to that of specialized models under few-shot or lightweight model conditions, indicating that the embedding has effectively internalized multi-source structural information (Fang et al., 2025; Houriez et al., 2025; Ma et al., 2026). This advantage makes rapid experimentation and few-shot mapping feasible, offering practical significance in data-scarce regions (Böröcz & Molnár, 2025; Fang et al., 2025; Ngo et al.). However, this benefit is primarily evident in scenarios with limited model complexity or insufficient sample size; under highly customized deep network frameworks, its performance gain is not consistently superior (Khan & Ahmad, 2025).

5.1.3. Cross-Regional Transfer

The advantages of AEF in cross-regional transfer stem from its highly standardized embedded structure and multimodal information fusion capabilities. By maintaining consistent feature representation across regions through unified resolution and standardized embedding, the model enables seamless transfer under diverse geographic and environmental conditions. In areas with frequent cloud cover or complex observation conditions, it demonstrates strong robustness to shadow noise from clouds (Seydi, 2025), reducing uncertainties caused by resampling errors and sensor differences (Cheng et al., 2026). Comparative experiments conducted in Global South cities and multiple regions have shown that AEF exhibits cross-regional consistency, making it particularly suitable for monitoring data-scarce or ecologically heterogeneous areas, while demonstrating stable spatial expression and strong noise resistance in disturbed zones and agricultural margins (Alvarez et al., 2025; Ryan et al., 2026). Additionally, the embedding space supports similarity measurement based on cosine distance, providing an out-of-the-box implementation pathway for regional matching (Böröcz & Molnár, 2025; Qu et al., 2026).

5.1.4. Environmental Foundation

Within the structure-process coupling modeling framework, AEF provides a unified static environmental semantic foundation for complex system modeling, capable of supplementing diverse environmental information to effectively support the integration and collaborative operation of multiple mechanisms. On one hand, it tightly integrates environmental semantics, spatial characteristics, and dynamic processes, enabling enhanced flexibility through rapid adjustments to model parameters and mechanism descriptions; on the other hand, AEF offers a solid basis for interdisciplinary collaboration, providing essential data support and methodological guarantees across fields such as socioeconomics (Pettersson & Daoud, 2025), environmental science (Ashfaq et al.), and public health research (Metz et al., 2025). These advantages are particularly evident in multi-mechanism modeling scenarios that require the integration of static background with dynamic driving mechanisms, thereby expanding the application scope of AEF to meet increasingly complex multidimensional modeling demands.

5.1.5. Scalable Application

As a precomputed, globally unified embedding asset, AEF offers high accessibility and reduces numerous steps in large-scale processing, ensuring broad application reachability and reproducibility (Lisaius et al., 2026; Metz et al., 2025; Zvonkov et al., 2025). It can easily accommodate data requirements of different scales, providing efficient computational support for both global modeling and localized detailed analysis. It efficiently supports collaborative processing across multiple years and regions, significantly lowering the threshold for cross-temporal and spatial applications (Kim et al., 2025; Ngo et al.). Moreover, its original high-resolution precision can adapt to various resolution requirements, effectively avoiding errors caused by resolution differences in traditional methods, thereby significantly improving data quality and accuracy (Fang et al., 2026).

5.2. Limitations

5.2.1. Time Resolution Constraints

AEF uses annual aggregation as its basic expression unit, which limits its ability to capture key dynamic changes or event-level variations within a year. For example, in seasonal phenological changes, the masking of climatic differences across different seasons within a year leads to the confusion between target signals and background variations in agricultural and fire monitoring (Fang et al., 2025; Seydi, 2025). In highly time-sensitive applications such as ozone pollution modeling, the inability to express instantaneous fluctuations in emissions and meteorology restricts the capacity to characterize peak values or extreme scenarios (Alvarez et al., 2025; Fang et al., 2026). In sudden events like flood mapping, the difficulty in reflecting short-term changes in environmental conditions such as land and vegetation during the period necessitates the introduction of additional dynamic driving variables or physical constraints for correction (Ashfaq et al.). Therefore, AEF is more suitable for tasks involving structural or long-term average state expressions, while it inherently lacks information in scenarios dominated by strong dynamic processes (Ma et al., 2026).

5.2.2. Lack of Semantic Interpretability

The embedding dimensions of AEF lack clear physical or biogeochemical semantics and cannot provide explicit mappings as traditional models do. Their representation mainly relies on feature importance ranking or posterior correlation analysis, making it difficult to directly map to specific environmental mechanisms (Esparza et al., 2025; Fang et al., 2025; Ngo et al.). Therefore, external variables or mutual information analysis are often required to further interpret the meaning of these embedding dimensions (Cheng et al., 2026; Qu et al., 2026). Some studies have shown that index methods with clear biophysical significance often perform better in certain ecological parameter inversion tasks, indicating that abstract representations embedded in some mechanism-driven scenarios are not always optimal (Lucero et al., 2026). In addition, some studies have pointed out that there may be redundancy or correlation structures among the embedding dimensions, and simple dimensionality reduction does not significantly improve model performance (Fang et al., 2026; Ryan et al., 2026).

5.2.3. Domain Shift

Although AEF provides a unified embedding space, its out-of-domain generalization capability significantly decreases during cross-domain transfer, particularly in regions with high ecological heterogeneity (Houriez et al., 2025; Ma et al., 2026). Cross-regional transfer depends not only on the similarity between the target and source regions but is also influenced by factors such as label sources, class definitions, and phenological regimes, which limit the model's generalization ability in new regions (Houriez et al., 2025; Murakami, 2025). Additionally, studies have found that transfer exhibits directionality; the effectiveness of transfer from certain regions is unidirectional, while reverse transfer may not yield equivalent results, indicating that embedding similarity does not equate to bidirectional transferability (Murakami, 2025). Furthermore, current generalization validation is limited to single-region or single-year experiments, and cross-regional stability has not been sufficiently tested (Khan & Ahmad, 2025; Lisaius et al., 2026; Ryan et al., 2026). Therefore, the cross-domain capability of AEF is conditional and cannot be simply assumed to enable seamless transfer across all regions.

5.2.4. Data Quality Constraints

As a feature representation tool, the performance ceiling of AEF is still constrained by the quality of ground truth data and observation errors. In modeling health and socioeconomic indicators, when data are missing, zero values are concentrated, or reporting biases are significant, embeddings struggle to surpass the upper limits imposed by data quality (Metz et al., 2025). In flood mapping and remote sensing inversion tasks, SAR threshold settings and observation errors similarly limit model performance (Ashfaq et al.). In large-scale spatial graph models, reliance on external point databases or gazetteer databases may introduce systematic biases, thereby affecting generalization performance (Pettersson & Daoud, 2025). Therefore, the potential for performance improvement in AEF remains constrained by objective data conditions.

5.2.5. Extreme Scenario Capability

The AEF embedding aims for a unified low-dimensional representation, where spatial semantics may tend to be smoothed, limiting performance in fine-grained differentiation or highly heterogeneous environments. In species-level or crop-type classification tasks, similar classes cluster closely in the embedding space, resulting in insufficient discriminability (Böröcz & Molnár, 2025; Lisaius et al., 2026). In complex ecological settings such as mountainous regions or ecotones, specialized indices that carry clear biophysical meanings perform more robustly, whereas AEF shows no significant advantage and even exhibits performance degradation (Lucero et al., 2026). In some comparative experiments, its performance also falls short of that of deep models specifically optimized for the task (Zvonkov et al., 2025). Therefore, in extreme or customized task scenarios, AEF is not consistently superior to dedicated models.

5.3. Dimension Interpretability Exploration

The original paper did not explicitly elaborate on the semantics of each dimension in the 64-dimensional embedding or its representation generation mechanism, nor did it establish direct connections with specific physical or ecological processes. Therefore, subsequent studies explored the interpretability of this model from multiple perspectives. The existing literature on interpretation methods can be roughly categorized into five paradigms.

5.3.1. Error Attribution

This paradigm focuses on answering “under what conditions the model performs well or poorly,” attributing performance variations to external factors such as domain shift, label systems, or task settings. Common approaches include decomposing overall metrics into diagnosable sub-metrics or revealing the causes of differences through comparative experiments. For example, in crop transfer studies, decomposing UA and PA per crop can concretize errors into “false alarm dominant” and “miss detection dominant,” making transfer differences clearer rather than relying solely on overall metrics like OA or Macro-F1 (Murakami, 2025). In cross-domain discussions, some studies attribute performance degradation to domain shifts caused by ecological heterogeneity and point out that label bias may lead models to appear ineffective, actually reflecting annotation inconsistency or changes in classification granularity (Houriez et al., 2025). Additionally, other studies emphasize that performance fluctuations stem from experimental design differences rather than the inherent merits or flaws of embeddings, through cross-year comparisons or contrasts with traditional features (Ma et al., 2026).

5.3.2. Statistical Attribution

This paradigm aims to answer the question of “which dimensions contribute practically to the task,” employing statistical tools such as SHAP, permutation importance, and gain importance to shift the focus of interpretation from “task outcomes” to “feature contributions.” In air pollution regression tasks, SHAP revealed high contributions from a few dimensions, highlighting the phenomenon of “importance sparsity” and supporting the view that data can be “reduced in dimensionality and compressed” (Alvarez et al., 2025). In spatial regression, combining permutation importance with PCA, it analyzed the relationship between important dimensions and error, aiding in determining whether spatially correlated factors were omitted (Fang et al., 2025). In water quality estimation, gain importance identified multiple water quality indicators sharing highly contributive dimensions, proposing the concept of a “reusable subset of watershed information” (Kim et al., 2025).

5.3.3. Mechanism Comparison

When the research problem itself exhibits structural characteristics (spatial dependence, propagation processes, or multi-source complementarity), some studies shift the focus of interpretation to “model structure and information channels”. The core idea of this paradigm is: rather than forcibly interpreting each of the 64 dimensions, it demonstrates through ablation experiments and structural comparisons which mechanism modules contribute to performance gains. For example, in transnational poverty mapping, by ablating factors such as “graph structure”, “unlabeled nodes”, and “fuzzy labels”, it was found that performance improvements mainly stem from spatial dependency modeling and training setup, rather than merely increasing feature dimensions (Pettersson & Daoud, 2025). In hydrological regionalization tasks, analyzing the size of donor neighbors and mutual information explains the mechanism of information transfer between similar watersheds (Qu et al., 2026). In wildfire spread risk modeling, by analyzing attention weights, graph centrality, and fusion coefficients, the mechanisms of propagation logic are revealed, rather than relying on a single-dimensional semantics of AEF (Esparza et al., 2025).

5.3.4. External Audit

In high-risk fields such as disaster and hydrology, the key to explainability is whether "the output conforms to basic physical common sense." Therefore, some studies incorporate physical constraints and diagnostic indicators to shift explanations from "describing models" to "constraining models," forming an explicit and auditable explanation framework. For example, in flood risk mapping, rules such as "increased rainfall should not decrease risk" and "high terrain should not show water accumulation artifacts" are explicitly encoded through monotonicity constraints and topographic consistency checks, thereby preventing models from fitting data while ignoring physical reality. The explanation chain then becomes "why does the model make this prediction — because it is constrained by rules" (Ashfaq et al.). In ecological and disturbance zone mapping, spatial structure metrics (e.g., edge density, local entropy, patch area, and other landscape pattern indices) quantify the "ecological realism" of maps, making model explanations more objective and aligning with both statistical patterns and fundamental ecological principles, thus enhancing credibility (Ryan et al., 2026).

5.3.5. Separability Test

This paradigm is closer to the perspective of representation learning. Studies commonly use Uniform Manifold Approximation and Projection (UMAP), clustering, and comparisons with external products for sanity checks to determine whether embeddings have learned meaningful semantic structures. By examining the geometric distribution of the embedding space and the separability between classes, one can assess whether the model has successfully learned effective representations. For example, studies compare unsupervised clustering results with products such as WorldCover to visually verify whether the embedded clusters correspond to major surface semantic types, thereby demonstrating that the embeddings are not merely noise compression (Zvonkov et al., 2025). In crop and land cover tasks, UMAP illustrates the separability of classes in the embedding space and interprets "catastrophic forgetting" as a collapse or aliasing of the embedding geometry in the target domain, attributing failure to the structure of the representation space rather than the classifier itself (Lisaius et al., 2026). Additionally, the discriminative approach combining prototype vectors with cosine distance offers a transparent decision process, enabling the question "why classified into this class?" to be directly explained by distance thresholds, providing interpretability at the methodological level (Böröcz & Molnár, 2025).

5.4. Future Work Prospects

5.4.1. Dynamic Expression Capability

The current AEF has limited capacity to characterize short-term dynamic changes, particularly showing deficiencies in expressing event-level or annual process variations, which is a key reason for its constrained performance in tasks related to ozone, flooding, agricultural phenology, and extreme peak events. First, the approach should transition from "annual static expression" to "multi-temporal granularity characterization", incorporating monthly, seasonal, or key phenological window embeddings under controlled computational costs to mitigate the smoothing effects of annual composites on short-term processes (Alvarez et al., 2025; Fang et al., 2025; Ma et al., 2026). Second, a temporal sampling consistency mechanism should be established to ensure stable embeddings for the same land surface unit across different temporal subsets, thereby enhancing robustness against sparse and non-uniform time series (Ryan et al., 2026; Seydi, 2025). Third, in hydrological and ecological contexts, integrating multi-year sequence modeling, object-level change detection, and inter-annual drift calibration can further improve the representation of process evolution (Cheng et al., 2026; Ngo et al.; Ryan et al., 2026). Overall, the goal is to extend AEF from an "annual semantic asset" into a "temporally expressive representational framework" with dynamic expression capabilities.

5.4.2. Cross-Domain Transfer Capability

The AEF exhibits performance degradation and directional bias during cross-ecological zone and cross-regime transfers, indicating insufficient cross-domain adaptability. Future improvement directions include several aspects. First, introducing domain adaptation and embedding alignment mechanisms can enhance generalization through multi-region joint training or fine-tuning with limited target domain data (Ma et al., 2026; Murakami, 2025). Second, for high temporal granularity data, it is important to construct an adaptation framework that accounts for both spatial and temporal domain shifts, such as seasonal phase differences and observational regime differences (Lisaius et al., 2026; Murakami, 2025). Third, in spatial modeling, adaptive donor sets can be built using similarity thresholds and diversity constraints to mitigate interference caused by neighborhood heterogeneity (Qu et al., 2026). Finally, adopting rigorous spatial cross-validation and transcontinental system evaluations is essential for establishing verifiable extrapolation evidence chains (Kim et al., 2025; Metz et al., 2025; Qu et al., 2026; Ryan et al., 2026). Transfer capability should be validated under institutionalized protocols rather than relying on single-region experiments.

5.4.3. Interpretability & Multi-Source Fusion

AEF currently lacks in embedding semantic decoding and sensitivity to specific process variables, necessitating enhanced interpretability of embeddings and improved coupling capabilities between multi-source fusion and structure-process interactions. First, it should be integrated with key mechanism variables relevant to specific applications. For example, high-resolution rainfall and soil moisture information can be introduced in flood and disaster risk modeling (Ashfaq et al.); intervention and supply chain variables can be incorporated in health and socioeconomic studies (Metz et al., 2025), among others. Furthermore, reconstructable or regressive constraints could be introduced during the training phase, combined with multi-task alignment strategies, to endow embeddings with clearer structures at both semantic and physical levels. Additionally, as embeddings gradually become "data assets," there should be a shift from "downstream variable stitching" toward a joint design of "upstream training objectives and release levels," ensuring that embeddings simultaneously satisfy requirements for critical physical signal reconstruction, clear semantic structure, and stable reliability under multi-task supervision.

5.4.4. Systematic Evaluation & Annotation

The performance ceiling of AEF is constrained by label quality and evaluation protocols. Future improvements should focus on refining the label generation process and class restructuring strategies to reduce confusion among similar classes (Houriez et al., 2025), establishing a cross-year consistent annotation system to mitigate label drift (Lisaius et al., 2026), and systematically defining probability thresholds and cost-sensitive decision rules in disaster-related tasks (Zvonkov et al., 2025). Additionally, more rigorous spatial cross-validation and multi-regional expansion validation methods should be adopted (Kim et al., 2025; Metz et al., 2025; Qu et al., 2026). If AEF gradually evolves into an embedded asset with long-term updates, the evaluation framework must also incorporate long-term consistency (cross-year drift stability), engineering usability (including compression and inference costs), and task-agnostic representation quality checks (e.g., distribution stability, reconstructability). In summary, performance evaluation must be grounded in a reproducible and transferable protocol framework to ensure that embedded assets can consistently deliver stable and reliable representational capabilities under varying temporal, spatial, and task conditions (Houriez et al., 2025; Kim et al., 2025; Lisaius et al., 2026; Metz et al., 2025).

5.4.5. Embedding Assetization and Lightweight Applications

In the future, AEF should evolve toward the concept of “embedding as a data asset,” enabling broader geospatial production and decision support. It is worth exploring lightweight interactive mapping based on precomputed embeddings, including point-selective sample labeling, prototype vector construction, and rapid similarity inference, allowing users without deep learning backgrounds to generate thematic maps or change layers with minimal annotations. At the same time, by integrating embedding subspace selection and lightweight classification head design, it is possible to reduce inference and storage costs while maintaining performance, thus extending embedding products to desktop-level and even mobile platforms. The core of this direction is to transform AEF from a “research model” into an “interactive, reproducible, and extensible spatial analysis infrastructure,” achieving widespread application of remote sensing intelligence (Houriez et al., 2025).

6. Conclusions

Against the backdrop of the rapid growth of Earth observation data and the increasing complexity of its applications, this study systematically reviews and comprehensively evaluates the AEF from the representational paradigm of “unified surface embedding.” Through a structured analysis of its conceptual positioning, data organization methods, and technical framework, this paper clarifies the paradigmatic characteristics of AEF as an analysis-ready representational data product rather than a parameterized foundational model. Based on systematic literature retrieval results, existing empirical studies are summarized in terms of research themes, regional distribution, methodological usage, and evaluation systems, revealing a research trajectory centered around core issues such as “plug-and-play usability, transfer boundaries, information structure, temporal sensitivity, information supplementation, and compensatory capabilities.” Integrating technical design with application feedback, it is evident that AEF offers significant advantages in reducing process complexity, compressing information, supporting cross-regional transfer, constructing a unified environmental foundation, and enabling large-scale applications. However, constraints remain regarding temporal resolution, semantic interpretability, domain shift, data quality, and extreme scenarios. Focusing on the issue of interpretability, this study further organizes discussions into five aspects: error attribution, statistical attribution, mechanism comparison, explicit auditing, and separability testing; current research remains in an exploratory developmental stage. Overall, AEF represents a significant exploration in the transition within the field of Earth observation from task-driven models to representation-driven data infrastructure. Its future development requires continued progress in dynamic expression, cross-domain adaptation, interpretability, multi-source fusion, evaluation and annotation, and asset-based applications to achieve a more robust, interpretable, and scalable global surface information representation system.

Funding: The work was supported by the National Natural Science Foundation of China [Grant number 452B2013].

Acknowledgments: This work was supported by the National Natural Science Foundation of China under Grant No. 452B2013, awarded to Shuyang Hou.

Conflicts of interest: The authors declare that they have no known competing financialinterests or personal relationships that could have appeared to influencethe work reported in this paper.

Appendix A: Surveyed Papers

Table A1. Surveyed papers and their corresponding metadata.

ID	Letter ID	Date	Title	DOI	Authors	Platform	Affiliation	Country
1	\	2025/7/29	AlphaEarth Foundations:An Embedding Field Model for Accurate and Efficient Global Mapping from Sparse Label Data	https://doi.org/10.48550/arXiv.2507.22291	Christopher F.Brown,Michal R.Kazmierski,Valerie J.Pasquarella,et al.	arXiv	Google DeepMind	USA
2	A	2025/8/15	Scalable Geospatial Data Generation Using AlphaEarth Foundations Model	https://doi.org/10.48550/arXiv.2508.11739	Luc Houriez,Sebastian Pilarski,Behzad Vahedi,et al.	arXiv	X,The Moonshot Factory;Bellwether;Stanford University	USA
3	B	2025/8/20	Within-and Cross-Regional Crop Classification for Cool Climate Upland Agriculture Using AlphaEarth	https://www.cabidigitallibrary.org/doi/abs/10.31220/agriRxiv.2025.00354	Keach Murakami	agriRxiv	Hokkaido Agricultural Research Center,NARO	Japan
4	C	2025/9/9	Deep Learning-Based Burned Area Mapping Using Bi-Temporal Siamese Networks and AlphaEarth Foundation Datasets	https://doi.org/10.48550/arXiv.2509.07852	Seyd Teymoor Seydi	arXiv	University of Tehran	Iran
5	D	2025/9/19	Leveraging AlphaEarth Foundations Embeddings for High-Accuracy County-Scale Corn and Soybean Yield Estimation	https://www.techrxiv.org/doi/full/10.36227/techrxiv.175825526.60198358	Jichao Fang,Mingda Wu,Zhou Zhang,et al.	TechRxiv	Northern Illinois University	USA
6	E	2025/10/10	Beyond AlphaEarth:Toward Human-Centered Spatial Representation via POI-Guided Contrastive Learning	https://doi.org/10.48550/arXiv.2510.09894	Junyuan Liu,Quan Qin,Guangsheng Dong,et al.	arXiv	University College London	UK
7	F	2025/10/17	Machine Learning for Urban Air Quality Prediction Using Google AlphaEarth Foundations Satellite Embeddings	https://doi.org/10.3390/rs17203472	Cesar Ivan Alvarez,Carlos Andrés Ulloa Vaca,Neptali Armando Echeverria Llumipanta	Remote Sensing	University of Augsburg	Germany
8	G	2025/10/29	Application and Validation of Geospatial Foundation Model Data for the Prediction of Health Facility Programmatic Outputs—A Case Study in Malawi	https://doi.org/10.48550/arXiv.2510.25954	Lynn Metz,Rachel Haggard,Michael Moszczynski,et al.	arXiv	Cooper/Smith	USA
9	H	2025/10/31	Estimating River Water Quality Indicators in South	https://doi.org/10.7780/kjrs.2025.41.5.10	Young Jun Kim,Gibeom Nam,Euiho Hwang	Korean Journal of	K-water Research Institute	South Korea

			Korea Using AlphaEarth Embeddings			Remote Sensing		
10	I	2025/11/3	Leveraging Compact Satellite Embeddings and Graph Neural Networks for Large-Scale Poverty Mapping	https://doi.org/10.48550/arXiv.2511.01408	Markus B.Pettersson,Adel Daoud	arXiv	Chalmers University of Technology	Sweden
11	J	2025/11/4	Cropland Mapping Using Geospatial Embeddings	https://doi.org/10.48550/arXiv.2511.02923	Ivan Zvonkov,Gabriel Tseng,Inbal Becker-Reshef,et al.	arXiv	University of Maryland,College Park	USA
12	K	2025/11/13	First Experiences with the AlphaEarth Foundations Model:A Cosine Distance-Based Evaluation	https://doi.org/10.12700/AIS.2025.008	Balázs Böröcz,Gábor Molnár	AIS 2025 Proceedings	Óbuda University	Hungary
13	L	2025/11/27	Evaluating AlphaEarth Foundation Embeddings for Pixel-and Object-Based Land Cover Classification in Google Earth Engine	https://doi.org/10.20944/preprints202511.2172.v1	Hayat Khan,Aftab Ahmad	Preprints.org	Board of Revenue,Khyber Pakhtunkhwa	Pakistan
14	M	2025/12/4	Performance Assessment and Application of AlphaEarth Embedding in Air Quality:Case of PM$_{2.5}$ in Changchun	https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5853943	Chunseng Fang,Benjamin M.Buhendwa,Farhana Riaz,et al.	SSRN	Jilin University	China
15	N	2025/12/7	Theory-Guided Deep Learning with AlphaEarth Embeddings for Flash Flood Prediction in Data-Scarce Regions	(Conference PDF)	Hassan Ashfaq,Muhammad Arsal,Anas Ashfaq	NeurIPS 2025	Ghulam Ishaq Khan Institute	Pakistan
16	O	2025/12/23	GraphFire-X:Physics-Informed Graph Attention Networks for Wildfire Preparedness	https://doi.org/10.48550/arXiv.2512.20813	Miguel Esparza,Vamshi Battal,Ali Mostafavi	arXiv	Texas A&M University	USA
17	P	2025/12/28	Forest canopy height map in Cuc Phuong National Park using field data and AI-extracted features AlphaEarth Foundations (AEF)	https://doi.org/10.15625/vap.2025.0212	Duc Anh Ngo,Anh Tuan Vu,Viet Luong Nguyen,et al.	EME 2025 Conference	Vietnam National Space Center	Vietnam
18	Q	2025/12/30	Harvesting Harvesting AlphaEarth: Benchmarking the geospatial foundation model for agricultural downstream tasks	https://doi.org/10.48550/arXiv.2601.00857	Yuchi Ma,Yawen Shen,Anu Swatantran,et al.	arXiv	Stanford University	USA
19	R	2026/1/4	Utilizing Earth Foundation Models to Enhance the Simulation Performance of Hydrological Models	https://doi.org/10.48550/arXiv.2601.01558	Pengfei Qu,Wenyu Ouyang,Chi Zhang,et al.	arXiv	Dalian University of Technology	China

			with AlphaEarth Embeddings					
20	S	2026/1/7	Spectral indices outperform AlphaEarth foundation embeddings for aboveground biomass estimation in tropical Andean Forests	https://doi.org/10.31223/X58X7V	Juan Camilo Rojas Lucero,Nicholas Kolarik,Jodi Brandt,et al.	EarthArXiv	Boise State University	USA
21	T	2026/1/12	From Landslide Conditioning Factors to Satellite Embeddings: Evaluating the Utilisation of Google AlphaEarth for Landslide Susceptibility Mapping using Deep Learning	https://doi.org/10.48550/arXiv.2601.07268	Yusen Cheng,Qinfeng Zhu,Lei Fan	arXiv	Xi'an Jiaotong-Liverpool University	China
22	U	2026/1/15	Streamlining Wetland Vegetation Mapping with AlphaEarth Embeddings: Comparable Accuracy to Traditional Methods with Cleaner Maps and Minimal Preprocessing	https://doi.org/10.3390/rs18020293	Shawn Ryan,Megan Powell,Joanne Ling,et al.	Remote Sensing	Department of Climate Change,Energy,the Environment and Water	Australia
23	V	2026/1/23	Embedding-Based Crop Type Classification in the Groundnut Basin of Senegal	https://doi.org/10.48550/arXiv.2601.16900	Madeline C.Lisaius,Srinivasan Keshav,Andrew Blake,et al.	arXiv	University of Cambridge	UK

References

Al-Emadi, S. A., Yang, Y., & Ofli, F. (2025). Analysing satellite imagery classification under spatial domain shift across geographic regions. *International Journal of Computer Vision*, *133*(11), 7672-7709.

Alvarez, C. I., Ulloa Vaca, C. A., & Echeverria Llumipanta, N. A. (2025). Machine learning for urban air quality prediction using Google AlphaEarth Foundations satellite embeddings: a case study of Quito, Ecuador. *Remote Sensing*, *17*(20), 3472.

Anderson, K., Ryan, B., Sonntag, W., Kavvada, A., & Friedl, L. (2017). Earth observation in service of the 2030 Agenda for Sustainable Development. *Geo-spatial Information Science*, *20*(2), 77-96.

Ashfaq, H., Arsal, M., & Ashfaq, A. Theory-Guided Deep Learning with AlphaEarth Embeddings for Flash Flood Prediction in Data-Scarce Regions.

Balsamo, G., Agusti-Parareda, A., Albergel, C., Arduini, G., Beljaars, A., Bidlot, J., Blyth, E., Bousserez, N., Boussetta, S., & Brown, A. (2018). Satellite and in situ observations for advancing global Earth surface modelling: A review. *Remote Sensing*, *10*(12), 2038.

Böröcz, B., & Molnár, G. (2025). First Experiences with the AlphaEarth Foundations Model: A Cosine Distance-Based Evaluation.

Brown, C. F., Kazmierski, M. R., Pasquarella, V. J., Rucklidge, W. J., Samsikova, M., Zhang, C., Shelhamer, E., Lahera, E., Wiles, O., & Ilyushchenko, S. (2025). Alphaearth foundations: An embedding field model for accurate and efficient global mapping from sparse label data. *arXiv preprint arXiv:2507.22291*.

Cheng, Y., Zhu, Q., & Fan, L. (2026). From Landslide Conditioning Factors to Satellite Embeddings: Evaluating the Utilisation of Google AlphaEarth for Landslide Susceptibility Mapping using Deep Learning. *arXiv preprint arXiv:2601.07268*.

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

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

Constantin, A., Fauvel, M., & Girard, S. (2022). Mixture of multivariate Gaussian processes for classification of irregularly sampled satellite image time-series. *Statistics and Computing*, *32*(5), 79.

Dong, H., Ma, W., Wu, Y., Zhang, J., & Jiao, L. (2020). Self-supervised representation learning for remote sensing image change detection based on temporal prediction. *Remote Sensing*, *12*(11), 1868.

Esparza, M., Battal, V., & Mostafavi, A. (2025). GraphFire-X: Physics-Informed Graph Attention Networks and Structural Gradient Boosting for Building-Scale Wildfire Preparedness at the Wildland-Urban Interface. *arXiv preprint arXiv:2512.20813*.

Fang, C., Buhendwa, B. M., Riaz, F., & Wang, J. (2026). Performance Assessment and Application of AlphaEarth Embedding In Air Quality-Case of PM2. 5 in Changchun. *Remote Sensing Applications: Society and Environment*, 101989.

Fang, J., Wu, M., Zhang, Z., & Luo, W. (2025). Leveraging AlphaEarth Foundations Embeddings for High-Accuracy County-Scale Corn and Soybean Yield Estimation. *Authorea Preprints*.

Feng, Z., Atzberger, C., Jaffer, S., Knezevic, J., Sormunen, S., Young, R., Lisaius, M. C., Immitzer, M., Jackson, T., & Ball, J. (2025). Tessera: Temporal embeddings of surface spectra for earth representation and analysis. *arXiv preprint arXiv:2506.20380*.

Gao, P., Geng, S., Zhang, R., Ma, T., Fang, R., Zhang, Y., Li, H., & Qiao, Y. (2024). Clip-adapter: Better vision-language models with feature adapters. *International Journal of Computer Vision*, *132*(2), 581-595.

Gong, X., Li, T., Wang, R., Hu, S., & Yuan, S. (2025). Beyond the remote sensing ecological index: A comprehensive ecological quality evaluation using a Deep-Learning-Based remote sensing ecological index. *Remote Sensing*, *17*(3), 558.

Guo, X., Lao, J., Dang, B., Zhang, Y., Yu, L., Ru, L., Zhong, L., Huang, Z., Wu, K., & Hu, D. (2024). Skysense: A multi-modal remote sensing foundation model towards universal interpretation for earth observation imagery. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,

Hafner, M., Katsantoni, M., Köster, T., Marks, J., Mukherjee, J., Staiger, D., Ule, J., & Zavolan, M. (2021). CLIP and complementary methods. *Nature Reviews Methods Primers*, *1*(1), 20.

Haut, J. M., Paoletti, M. E., Moreno-Álvarez, S., Plaza, J., Rico-Gallego, J.-A., & Plaza, A. (2021). Distributed deep learning for remote sensing data interpretation. *Proceedings of the IEEE*, *109*(8), 1320-1349.

Houriez, L., Pilarski, S., Vahedi, B., Ahmadalipour, A., Scully, T. H., Aflitto, N., Andre, D., Jaffe, C., Wedner, M., & Mazzola, R. (2025). Scalable geospatial data generation using AlphaEarth foundations model. *arXiv preprint arXiv:2508.11739*.

Huang, Z., Yan, H., Zhan, Q., Yang, S., Zhang, M., Zhang, C., Lei, Y., Liu, Z., Liu, Q., & Wang, Y. (2025). A survey on remote sensing foundation models: From vision to multimodality. *arXiv preprint arXiv:2503.22081*.

Huo, C., Chen, K., Zhang, S., Wang, Z., Yan, H., Shen, J., Hong, Y., Qi, G., Fang, H., & Wang, Z. (2025). When remote sensing meets foundation model: A survey and beyond. *Remote Sensing*, *17*(2), 179.

Jean, N., Wang, S., Samar, A., Azzari, G., Lobell, D., & Ermon, S. (2019). Tile2vec: Unsupervised representation learning for spatially distributed data. Proceedings of the AAAI Conference on Artificial Intelligence,

Jin, C., Jiang, X., Wen, L., Wu, C., Xu, X., & Jiao, J. (2026). Assessing the Utility of Satellite Embedding Features for Biomass Prediction in Subtropical Forests with Machine Learning. *Remote Sensing*, *18*(3), 436.

Ju, J., Zhou, Q., Freitag, B., Roy, D. P., Zhang, H. K., Sridhar, M., Mandel, J., Arab, S., Schmidt, G., & Crawford, C. J. (2025). The Harmonized Landsat and Sentinel-2 version 2.0 surface reflectance dataset. *Remote sensing of environment*, *324*, 114723.

Kartal, H., Alganci, U., & Sertel, E. (2018). Automated orthorectification of VHR satellite images by SIFT-based RPC refinement. *ISPRS International Journal of Geo-Information*, *7*(6), 229.

Kavvada, A., Metternicht, G., Kerblat, F., Mudau, N., Haldorson, M., Laldaparsad, S., Friedl, L., Held, A., & Chuvieco, E. (2020). Towards delivering on the sustainable development goals using earth observations. In (Vol. 247, pp. 111930): Elsevier.

Khan, H., & Ahmad, A. (2025). Evaluating AlphaEarth Foundation Embeddings for Pixel-and Object-Based Land Cover Classification in Google Earth Engine.

Kim, Y. J., Nam, G., & Hwang, E. (2025). Estimating River Water Quality Indicators in South Korea Using AlphaEarth Embeddings. *Korean Journal of Remote Sensing*, *41*(5), 829-841.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., & Lo, W.-Y. (2023). Segment anything. Proceedings of the IEEE/CVF international conference on computer vision,

Koldasbayeva, D., Tregubova, P., Gasanov, M., Zaytsev, A., Petrovskaia, A., & Burnaev, E. (2024). Challenges in data-driven geospatial modeling for environmental research and practice. *Nature Communications*, *15*(1), 10700.

Li, D., Tong, Q., Li, R., Gong, J., & Zhang, L. (2012). Current issues in high-resolution earth observation technology. *Science China earth sciences*, *55*(7), 1043-1051.

Lisaius, M. C., Keshav, S., Blake, A., & Atzberger, C. (2026). Embedding-based Crop Type Classification in the Groundnut Basin of Senegal. *arXiv preprint arXiv:2601.16900*.

Liu, H., Zhang, H. K., Huang, B., Yan, L., Tran, K. K., Qiu, Y., Zhang, X., & Roy, D. P. (2024). Reconstruction of seamless harmonized Landsat Sentinel-2 (HLS) time series via self-supervised learning. *Remote sensing of environment*, *308*, 114191.

Liu, J., Qin, Q., Dong, G., Wang, X., Feng, J., Zeng, Z., & Cheng, T. (2025). Beyond AlphaEarth: toward human-centered spatial representation via POI-guided contrastive learning. *arXiv preprint arXiv:2510.09894*.

Lu, S., Guo, J., Zimmer-Dauphinee, J. R., Nieusma, J. M., Wang, X., VanValkenburgh, P., Wernke, S. A., & Huo, Y. (2025). Vision foundation models in remote sensing: A survey. *IEEE Geoscience and Remote Sensing Magazine*, *13*(3), 190-215.

Lucero, J. C. R., Kolarik, N., Brandt, J., Caughlin, T., & Solano, T. (2026). Spectral indices outperform AlphaEarth foundation embeddings for aboveground biomass estimation in tropical Andean Forests.

Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., & Johnson, B. A. (2019). Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS journal of photogrammetry and remote sensing*, *152*, 166-177.

Ma, Y., Shen, Y., Swatantran, A., & Lobell, D. B. (2026). Harvesting AlphaEarth: Benchmarking the geospatial foundation model for agricultural downstream tasks. *International Journal of Applied Earth Observation and Geoinformation*, *149*, 105258.

Manas, O., Lacoste, A., Giró-i-Nieto, X., Vazquez, D., & Rodriguez, P. (2021). Seasonal contrast: Unsupervised pre-training from uncurated remote sensing data. Proceedings of the IEEE/CVF international conference on computer vision,

Marsocci, V., & Audebert, N. (2025). Cross-sensor self-supervised training and alignment for remote sensing. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.

Metz, L., Haggard, R., Moszczynski, M., Asbah, S., Mwase, C., Khomani, P., Smith, T., Cooper, H., Mwale, A., & Muslim, A. (2025). Application and Validation of Geospatial Foundation Model Data for the Prediction of Health Facility Programmatic Outputs--A Case Study in Malawi. *arXiv preprint arXiv:2510.25954*.

Murakami, K. (2025). Within-and cross-regional crop classification for cool climate upland agriculture using AlphaEarth. *agriRxiv*(2025), 20250417408.

Ngo, D. A., Vu, A. T., Nguyen, V. L., Nguyen, T. C., Truong, N. K. T., & Nguyen, T. B. Forest canopy height map in Cuc Phuong National Park using field data and AI-extracted features AlphaEarth Foundations (AEF).

Pascual, A., & Guerra-Hernández, J. (2026). Integration of Google's Alpha Earth Foundations into biomass estimation combined with GEDI spaceborne lidar and field inventory data. *Forest Ecology and Management*, *606*, 123550.

Pelletier, C., Webb, G. I., & Petitjean, F. (2019). Temporal convolutional neural network for the classification of satellite image time series. *Remote Sensing*, *11*(5), 523.

Persello, C., Wegner, J. D., Hänsch, R., Tuia, D., Ghamisi, P., Koeva, M., & Camps-Valls, G. (2022). Deep learning and earth observation to support the sustainable development goals: Current approaches, open challenges, and future opportunities. *IEEE Geoscience and Remote Sensing Magazine*, *10*(2), 172-200.

Pettersson, M. B., & Daoud, A. (2025). Leveraging Compact Satellite Embeddings and Graph Neural Networks for Large-Scale Poverty Mapping. *arXiv preprint arXiv:2511.01408*.

doi:10.20944/preprints202605.0981.v1

Qu, P., Ouyang, W., Zhang, C., Chai, Y., Xu, S., Ye, L., Piao, Y., Zhang, M., & Lu, H. (2026). Utilizing Earth Foundation Models to Enhance the Simulation Performance of Hydrological Models with AlphaEarth Embeddings. *arXiv preprint arXiv:2601.01558*.

Reinartz, P., Müller, R., Schwind, P., Suri, S., & Bamler, R. (2011). Orthorectification of VHR optical satellite data exploiting the geometric accuracy of TerraSAR-X data. *ISPRS journal of photogrammetry and remote sensing*, *66*(1), 124-132.

Ryan, S., Powell, M., Ling, J., & Wen, L. (2026). Streamlining Wetland Vegetation Mapping with AlphaEarth Embeddings: Comparable Accuracy to Traditional Methods with Cleaner Maps and Minimal Preprocessing. *Remote Sensing*, *18*(2).

Seydi, S. T. (2025). Deep Learning-Based Burned Area Mapping Using Bi-Temporal Siamese Networks and AlphaEarth Foundation Datasets. *arXiv preprint arXiv:2509.07852*.

Solórzano, J. V., Mas, J. F., Gallardo-Cruz, J. A., Gao, Y., & de Oca, A. F.-M. (2023). Deforestation detection using a spatio-temporal deep learning approach with synthetic aperture radar and multispectral images. *ISPRS journal of photogrammetry and remote sensing*, *199*, 87-101.

Szwarcman, D., Roy, S., Fraccaro, P., Gíslason, O. E., Blumenstiel, B., Ghosal, R., De Oliveira, P. H., de Sousa Almeida, J. L., Sedona, R., & Kang, Y. (2025). Prithvi-eo-2.0: A versatile multi-temporal foundation model for earth observation applications. *IEEE Transactions on Geoscience and Remote Sensing*.

Tseng, G., Cartuyvels, R., Zvonkov, I., Purohit, M., Rolnick, D., & Kerner, H. (2023). Lightweight, pre-trained transformers for remote sensing timeseries. *arXiv preprint arXiv:2304.14065*.

Vermote, E., Justice, C., Claverie, M., & Franch, B. (2016). Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. *Remote sensing of environment*, *185*, 46-56.

Wang, H., Mao, K., Yuan, Z., Shi, J., Cao, M., Qin, Z., Duan, S., & Tang, B. (2021). A method for land surface temperature retrieval based on model-data-knowledge-driven and deep learning. *Remote sensing of environment*, *265*, 112665.

Wu, B., Zhang, M., Zeng, H., Tian, F., Potgieter, A. B., Qin, X., Yan, N., Chang, S., Zhao, Y., & Dong, Q. (2023). Challenges and opportunities in remote sensing-based crop monitoring: A review. *National Science Review*, *10*(4), nwac290.

Wu, H., Jiao, H., Hou, S., Qing, Y., Xu, Q., Liang, J., Zhang, X., Gui, Z., Guan, X., & Xiang, L. (2025). GeoRRDI: An explainable and multi-source data-driven framework for rural revitalization assessment with spatial heterogeneity consideration. *Ecological Indicators*, *181*, 114387.

Xiao, A., Xuan, W., Wang, J., Huang, J., Tao, D., Lu, S., & Yokoya, N. (2025). Foundation models for remote sensing and earth observation: A survey. *IEEE Geoscience and Remote Sensing Magazine*.

Xu, C., Du, X., Fan, X., Giuliani, G., Hu, Z., Wang, W., Liu, J., Wang, T., Yan, Z., & Zhu, J. (2022). Cloud-based storage and computing for remote sensing big data: a technical review. *International Journal of Digital Earth*, *15*(1), 1417-1445.

Xu, Y., Ma, Y., & Zhang, Z. (2024). Self-supervised pre-training for large-scale crop mapping using Sentinel-2 time series. *ISPRS journal of photogrammetry and remote sensing*, *207*, 312-325.

Yamamoto, T., Akahoshi, H., & Kitazawa, S. (2025). Emergence of human-like attention and distinct head clusters in self-supervised vision transformers: A comparative eye-tracking study. *Neural Networks*, *189*, 107595.

Yao, X., Li, G., Xia, J., Ben, J., Cao, Q., Zhao, L., Ma, Y., Zhang, L., & Zhu, D. (2019). Enabling the big earth observation data via cloud computing and DGGS: Opportunities and challenges. *Remote Sensing*, *12*(1), 62.

Zeng, J., Gu, Y., Qin, C., Jia, X., Deng, S., Xu, J., & Tian, H. (2024). Unsupervised domain adaptation for remote sensing semantic segmentation with the 2D discrete wavelet transform. *Scientific Reports*, *14*(1), 23552.

Zeng, W., Cheng, M., Yuan, Z., Dai, W., Wu, Y., Liu, W., & Wang, C. (2024). Domain adaptive remote sensing image semantic segmentation with prototype guidance. *Neurocomputing*, *580*, 127484.

Zhang, H., Li, F., Liu, S., Zhang, L., Su, H., Zhu, J., Ni, L. M., & Shum, H.-Y. (2022). Dino: Detr with improved denoising anchor boxes for end-to-end object detection. *arXiv preprint arXiv:2203.03605*.

Zhao, Y., Jiao, H., Fei, T., Huang, J., & You, N. (2026). Spatial distribution prediction and scale effect analysis of urban daytime noise based on remote sensing images: a case study of Chengdu. *Journal of Environmental Management*, *397*, 128379.

Zhou, Y. n., Wang, Y., Yan, N. n., Feng, L., Chen, Y., Wu, T., Gao, J., Zhang, X., & Zhu, W. (2023). Contrastive-learning-based time-series feature representation for parcel-based crop mapping using incomplete sentinel-2 image sequences. *Remote Sensing*, *15*(20), 5009.

Zhu, S., Wu, C., Du, B., & Zhang, L. (2023). Style and content separation network for remote sensing image cross-scene generalization. *ISPRS journal of photogrammetry and remote sensing*, *201*, 1-11.

Zhu, Z., & Woodcock, C. E. (2012). Object-based cloud and cloud shadow detection in Landsat imagery. *Remote sensing of environment*, *118*, 83-94.

Zvonkov, I., Tseng, G., Becker-Reshef, I., & Kerner, H. (2025). Cropland Mapping using Geospatial Embeddings. *arXiv preprint arXiv:2511.02923*.