

Foundation models in Earth system science: Advances, challenges, and future paths

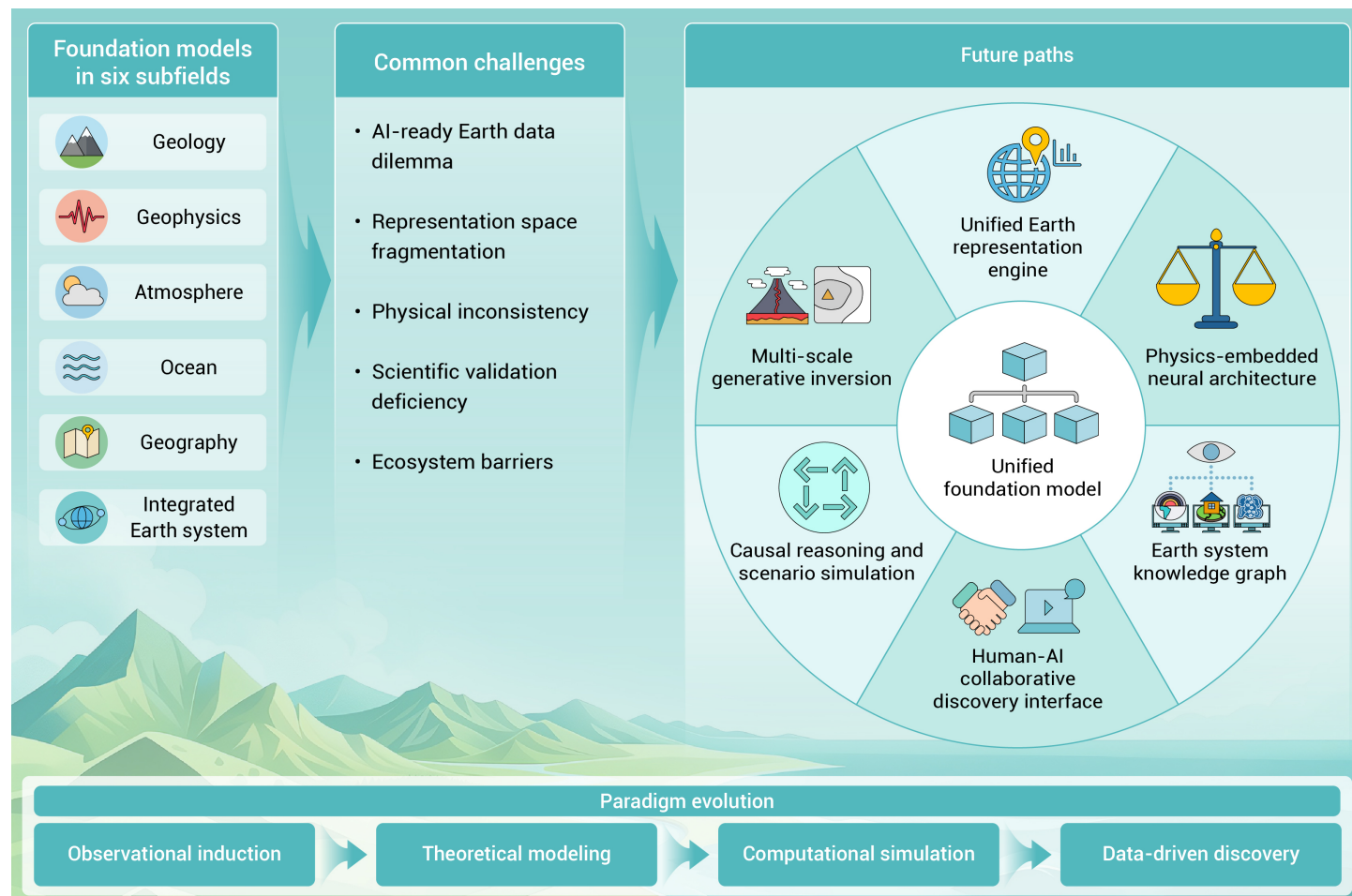
Fenzhen Su,^{1,2,13,*} Yinxia Cao,^{1,2,13} Xin Li,³ Haoyu Wang,^{1,2} Rong Fan,^{1,2} Shuyuan Huang,^{1,2} Wenxing He,^{1,2} Ziyang Chen,^{1,2} Peng Zhang,^{1,2} Liang Zhao,⁴ Tinghua Ai,⁵ Gang Huang,⁶ Haosu Tang,⁷ Xiaofeng Li,⁸ Zimeng Zhao,⁸ Changming Dong,⁹ Zhenhong Li,^{10,11} Libo Wang,¹² Junjie Wang,^{1,2} and Chengyuan Ju^{1,2}

*Correspondence: sufz@lreis.ac.cn

Received: April 14, 2026; Accepted: September 15, 2026; Published Online: September 20, 2026; <https://doi.org/10.59717/j.tirv.2026.100014>

© 2026 The Author(s). This is an open access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>).

GRAPHICAL ABSTRACT



PUBLIC SUMMARY

- This review examines foundation models across six fields of Earth system science.
- Limited access to usable data and difficulties in respecting physical laws remain major challenges.
- Future models could connect information across Earth's systems and incorporate physical knowledge.
- These advances could improve planetary monitoring and simulations of future Earth system changes.

Foundation models in Earth system science: Advances, challenges, and future paths

Fenzhen Su,^{1,2,13,*} Yinxia Cao,^{1,2,13} Xin Li,³ Haoyu Wang,^{1,2} Rong Fan,^{1,2} Shuyuan Huang,^{1,2} Wenxing He,^{1,2} Ziyang Chen,^{1,2} Peng Zhang,^{1,2} Liang Zhao,⁴ Tinghua Ai,⁵ Gang Huang,⁶ Haosu Tang,⁷ Xiaofeng Li,⁸ Zimeng Zhao,⁸ Changming Dong,⁹ Zhenhong Li,^{10,11} Libo Wang,¹² Junjie Wang,^{1,2} and Chengyuan Ju,^{1,2}

¹State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China

²College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 101408, China

³National Tibetan Plateau Data Center, State Key Laboratory of Tibetan Plateau Earth System, Environment and Resources, Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Beijing 100101, China

⁴Key Laboratory of Deep Petroleum Intelligent Exploration and Development, Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing 100029, China

⁵School of Resource and Environmental Sciences, Wuhan University, Wuhan 430079, China

⁶State Key Laboratory of Earth System Numerical Modeling and Application, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China

⁷School of Geography and Planning, University of Sheffield, Sheffield S3 7ND, UK

⁸Key Laboratory of Ocean Observation and Forecasting and Laboratory of Ocean Circulation and Waves, Institute of Oceanology, Chinese Academy of Sciences, Qingdao 266071, China

⁹School of Marine Sciences, Nanjing University of Information Science and Technology, Nanjing 210044, China

¹⁰State Key Laboratory of Loess Science, Chang'an University, Xi'an 710054, China

¹¹College of Geological Engineering and Geomatics, Chang'an University, Xi'an 710054, China

¹²School of Remote Sensing and Geomatics Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China

¹³These authors contributed equally

*Correspondence: sufz@reis.ac.cn

Received: April 14, 2026; Accepted: September 15, 2026; Published Online: September 20, 2026; <https://doi.org/10.59717/j.tirv.2026.100014>

© 2026 The Author(s). This is an open access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>).

Citation: Su F., Cao Y., Li X., et al. (2026). Foundation models in Earth system science: Advances, challenges, and future paths. *The Innovation Reviews* 1:100014.

The convergence of massive Earth observation datasets and emerging foundation models offers an opportunity to address the complex challenges in Earth System Science (ESS). However, the direct application of foundation models to ESS is often hindered by domain gaps, characterized by the need for physical consistency, the complexity of Earth's spherical geometries, and the non-stationarity of the multisphere coupled Earth systems. Building on prior reviews that prioritize model architecture or view geoscience as a monolithic entity, this study provides a domain-centric synthesis of foundation models across six distinct subfields, including Geology, Geophysics, Atmosphere, Ocean, Geography, and the Integrated Earth System. By analyzing the applications and limitations of Artificial Intelligence (AI) within these sub-fields, this work identifies unique domain barriers and cross-cutting challenges, including the AI-ready Earth data dilemma, representation space fragmentation, physical inconsistency, scientific validation deficiency, and ecosystem barriers. In this context, we provide future paths for the next generation of ESS foundation models, driven by unified Earth representation engine, a physics-embedded neural architecture, an Earth system knowledge graph, multi-scale generative inversion, causal reasoning, and human-AI collaborative discovery interface. In this way, foundation models can potentially complement traditional numerical models to monitor planetary status and simulate its future evolution, thereby equipping the scientific community with new tools for holistic planetary stewardship and complex decision-making.

INTRODUCTION

Earth System Science (ESS) is facing the scientific challenge of understanding the Earth as an interconnected and adaptive system¹ of multiple spheres, including the atmosphere, cryosphere, lithosphere, geography, ocean, biosphere, and anthroposphere. The demand for precise, real-time, and predictive insights into planetary dynamics has grown, driven by compounding environmental and anthropogenic challenges from climate change, biodiversity loss,² extreme weather events,³ and geopolitical crises.⁴ Meeting this demand has become increasingly feasible with the proliferation and accumulation of Big Earth Data from satellite constellations, Internet of Things (IoT) sensor networks, ground observation sites, reanalysis datasets,^{5,6} etc.

Following Jim Gray's seminal framework,⁷ ESS has evolved through four complementary methodological paradigms, i.e., observational induction,

theoretical modeling, computational simulation, and data-driven discovery. Contemporary Earth System Science does not follow a linear progression but rather integrates these paradigms into a synergistic ecosystem, where physics-based computational models coexist with and are enhanced by data-driven AI approaches to address complex planetary challenges. Rapid advances in natural language processing and computer vision have contributed to the development of AI models such as ChatGPT,⁸ DeepSeek,⁹ and Sora.¹⁰ The application of AI to scientific data, including remotely sensed observations of geoscientific phenomena, forms part of AI for Science (AI4Science) and supports an emerging research paradigm in geoscience. While effective, most domain-specific networks operated in silos, where the model trained to detect water bodies could not predict soil moisture, and similarly the model trained on optical imagery failed when applied to radar data. This chimney-style development paradigm requires expensive human-annotated labels and often has limited generalization ability across diverse scenarios.

Recent work has brought foundation models into Earth System Science (ESS).^{11,12} In this review, foundation models refer to models pretrained on broad and heterogeneous datasets that learn reusable representations or capabilities and can be adapted to multiple tasks or settings. We use three criteria when deciding whether a model falls within this scope: evidence of broad pretraining, capability beyond a single target task, and documented transfer or downstream adaptation. We distinguish these models from domain-specific pretrained models with a narrower transfer scope, task-oriented AI models developed for a defined prediction problem, hybrid physics-AI systems, and conventional physics-based models used as scientific references. Model size or forecast speed alone is not sufficient for classification as a foundation model. This distinction is especially important in ESS, where spherical geometry, multiscale interactions, non-stationarity, and conservation requirements can restrict transfer beyond the training distribution.¹³⁻¹⁶

The integration of AI with ESS combines complementary strengths: data-driven approaches extract patterns from high-dimensional observations, while physics-based models provide a mechanistic understanding of the underlying processes. This synergy is particularly transformative for foundation models, which uniquely bridge statistical learning and physical laws to generate not merely predictions but scientifically interpretable insights. Despite growing interest in this integration, current literature lacks a systematic synthesis of how foundation models specifically adapt to the heterogeneous physical regimes across Earth's subsystems. Existing surveys

predominantly address methodological frameworks and domain-specific applications.¹⁷⁻²⁰ For instance, Persello et al.¹⁹ examined deep learning within the specific context of Earth observation to advance Sustainable Development Goals. Zhao et al.²⁰ highlighted the shift toward hybrid and physics-informed paradigms to address data scarcity and interpretability. A closely related review examined geoscience foundation models with an emphasis on their architectures and remote-sensing applications,¹⁷ providing a basis for further synthesis of how their capabilities and limitations vary across Earth science domains.

This review examines foundation models across six interconnected fields: geology, geophysics, atmospheric science, ocean science, geography, and integrated Earth system modeling. These fields are compared using a common set of questions concerning data conditions, transferability, physical consistency, cross-sphere coupling, and scientific validation. The review also considers how foundation models relate to numerical modeling, data assimilation, reanalysis, and process-based methods. On this basis, five recurring limitations are connected to six research paths, with near-term technical priorities distinguished from longer-term proposals concerning causal simulation, human-AI collaboration, and Earth system governance. This structure provides an integrated assessment of where foundation models currently contribute to Earth system research and where further scientific evidence is still required.

PARADIGM EVOLUTION

Earth System Science has evolved through four successive paradigms, including observational induction, theoretical modeling, computational simulation, and data-driven discovery. Observational work establishes empirical facts, theoretical models formalize these facts with physical laws, computational simulation numerically solves complex physical systems, and data-driven learning uses big data and AI to accelerate understanding and prediction.

Observational induction

From the 18th to the late 19th century, Earth System Science was mainly dominated by empirical observation and inductive description.²¹ Through field mapping, sampling, and comparative analysis, researchers gradually reconstructed Earth's material composition, spatial structure, and evolutionary history. The epistemic style was fact-driven (grounded in direct, perceptible observations), classification-based (typologies of strata, landforms, minerals, climates), regionally focused (systematic documentation of spatial and geographical variation), and description-oriented (emphasizing depiction over prediction or calculation). Geology and geography were the core disciplines supported by advances in mining, navigation, field observations, stratigraphic mapping, and landform analysis, yielding early models of Earth's structure and evolution.^{22,23}

This paradigm prioritized the accumulation of empirical facts and the inductive summarization of regularities.²⁴ Representative achievements included the law of stratigraphic succession,^{25,26} early plate-tectonic hypotheses,²⁷ and geomorphological classification schemes.²⁸ Geological research extended from microscopic rock and mineral structures to regional tectonics and plate motions, and already acknowledged the complexity and multi-scale nature of geological systems.²⁹ Geography concentrated on surface morphology, climate, and human-environment relations, constructing an early Earth portrait through regional comparison and spatial classification. This phase established the empirical database and systematic classification framework underpinning later theoretical modeling. Its observational-inductive basis, however, produced a fragmented, locally constrained picture that lacked a unified Earth-system perspective and a physical explanatory framework.

Theoretical modeling and computational simulation

Since the 20th century, Earth System Science has entered the computational simulation paradigm, grounded in theoretical modeling based on physical laws and mathematical equations.³⁰ Conservation of energy, momentum, and mass provides the foundation for models formulated as partial differential equations and solved via finite difference and finite element methods. This has driven the development of relatively independent yet related model-

ing systems in geology, geophysics, atmospheric science, and oceanography, which are calibrated and validated against experimental and observational data.³¹

Breakthroughs in this paradigm converge in three areas. First, physical mechanisms have been quantitatively formalized. Key processes in geology, geophysics, and atmospheric-ocean systems are expressed as solvable equation sets, enabling direct mapping from mechanisms to mathematics.³² Geological modeling based on partial differential equations provides foundational treatments of deformation, fluid flow, and thermo-mechanical coupling.³³ In geophysics, governing equations and observation operators jointly support imaging and inversion with incomplete data. In atmospheric and ocean sciences, geophysical fluid dynamics underpins circulation models enabling reproducible multi-scale transport simulations. Second, modeling frameworks have been standardized and hierarchically organized from local to global scales. Microscale turbulence and boundary-layer processes link with mesoscale and planetary dynamics through modular interfaces. Global Climate Models (GCMs)³⁴ and Earth System Models (ESMs)³⁵ adopt dynamically adaptive grids, boundary conditions, and parameterization schemes. Remote sensing and Geographic Information Systems integration similarly yields unified geographic simulation workflows.³⁶ Third, cross-scale coupling has become central. Parameterization schemes project unresolved microscopic processes onto model grids, closing relationships between sub-grid physics and large-scale climate states, while also reflecting computational and numerical constraints. In geology and geophysics, equivalent media and effective constitutive laws typically encapsulate multiscale mechanics; in atmospheric and ocean modeling, turbulent mixing, cloud microphysics, and air-sea exchanges enter governing equations through empirical-semi-theoretical parameterizations for large-scale prediction.^{37,38}

Data-driven discovery

Since the beginning of the 21st century, the convergence of IoT, the Internet, and communication technologies has enabled real-time automated Earth observation, propelling ESS into an unprecedented data-intensive era.^{5,6} This evolution does not replace but rather complements and augments previous paradigms, which have always utilized data, yet operated within data-sparse regimes constrained by measurement limitations and computational barriers. In contrast, the contemporary data-intensive paradigm, powered by supercomputing and AI,³⁹ constructs computable knowledge systems through unified representation, cross-modal learning, and real-time assimilation, characterized by four transformative features. First, it transcends purely equation-driven approaches by using AI techniques to extract complex dynamic patterns, forming transferable statistical-physical representations from data-dense environments. Second, it evolves from fragmented local models toward unified Earth representations via large-scale multimodal foundation models⁴⁰⁻⁴² that align and compress cross-sphere features. Third, it advances from static deduction to intelligent feedback cycles by integrating data assimilation⁴³ with online learning for adaptive, near-real-time Earth system updates. Fourth, it bridges disciplinary fragmentation through transfer learning and multi-task training, enabling holistic sphere coordination and cross-domain knowledge flow previously unattainable in data-scarce paradigms.

Foundation models^{44,45} are increasingly positioned as key enablers of the data-driven paradigm, aiming to shift Earth system simulation from task-specific customization toward model unification. Their competencies can be grouped into five domains, each with verifiable examples across disciplines. First, multimodal fusion can combine images, text, time series, and volumetric data to jointly represent satellite, radar, and ground observations. In geosciences, large multimodal models can link literature, well logs, and core images.⁴⁴ Second, generative modeling and uncertainty quantification reconstruct sparse observations and missing data, using probabilistic outputs and ensemble methods to estimate confidence intervals and structural uncertainty. For example, AI has been used to reconstruct missing historical climate-extreme records from sparse observations,⁴⁶ producing gridded data with quantified uncertainty. Third, physics-informed learning⁴⁷ can embed conservation laws and constitutive relations into network structures or loss functions, improving stability, integrability, and long-term conservation. For example, it enforces momentum, energy, and mass conservation in atmo-

spheric modeling, while physics-guided extensions successfully integrate radiative transfer to enhance physical consistency. Fourth, cross-sphere unified modeling uses Transformers and graph neural networks to couple atmosphere, ocean, and land surface simulations. For example, the 3D-Geoformer model⁴⁸ successfully represents multivariate ocean-atmosphere interactions for El Niño-Southern Oscillation (ENSO) prediction. Fifth, self-supervised and causal learning⁴⁹ can extract universal representations from unlabeled or weakly labeled data using causal constraints.

Built upon century-long observational archives and physical theory, ESS faces an inherent challenge: its processes span spatial scales from microscopic mineral interactions to macroscopic planetary circulation patterns, and temporal scales from instantaneous chemical reactions to million-year tectonic cycles, exceeding the capacity of any single paradigm. Under conditions of strong nonlinearity, multiscale interactions, and incomplete mechanistic understanding, observational evidence remains fragmented, theoretical models rely on idealized assumptions, and computational simulations face a combinatorial explosion in parameter spaces. Emerging foundation models,^{44,45} through scaling laws, evolve non-linearly with data volume, parameter scale, and computational resources, showing promising results in unified representation of complex patterns, cross-domain generalization, multimodal information fusion, and spatiotemporal prediction. More importantly, ESS possesses uniquely favorable scaling conditions, e.g., systematically curated multi-decadal observations across atmospheric, ocean, and lithospheric domains, combined with physics-based regularization from conservation laws. Therefore, ESS foundation models constrained by physical consistency hold the potential to improve the accuracy of future evolution simulations and identify novel physical correlations within high-dimensional representation spaces.⁵⁰ This enables the identification of cross-sphere coupling mechanisms that are difficult to capture by traditional methods, thereby promoting a new paradigm that is predictable, interpretable, and intervention-oriented, while providing support for Earth system governance with both predictive capability and causal interpretability.

PROGRESS

The deep integration of AI into ESS is increasingly positioning foundation models as promising tools for cross-disciplinary research (Figure 1 and Table 1). ESS foundation models have primarily evolved into two complementary paradigms based on their data regimes and target tasks: 1) prediction-centric models, including GraphCast,⁵¹ Pangu-Weather,⁵² and FourCastNet,⁵³ which are trained on homogeneous gridded and reanalysis data to deliver precise spatiotemporal forecasts of atmospheric, oceanic, and land-surface dynamics; and 2) knowledge/literature-centric models, e.g., GeoBERT⁵⁴ and GeoGalactica,⁵⁵ which extract insights from unstructured geoscience literature to support document understanding, domain-specific question answering, and knowledge reasoning, thereby infusing predictive outcomes with domain expertise and interpretability. While current models face bottlenecks in cross-scale fusion and few-shot generalization, advances in multimodal data and algorithm design promise to reveal Earth system coupling mechanisms, enhance disaster warning and resource exploration capabilities, and accelerate the intelligent transformation of Earth science.

Geological foundation models

Geological foundation models deal with complex cross-scale, cross-modal data for characterizing uncertain subsurface systems. Their core tasks align with fundamental geological cognition and practical demands, focusing on four key directions. For paleontology and stratigraphic identification, they help reduce reliance on expert-dependent qualitative interpretation by mapping multimodal data (core images, logging curves, geological texts) into a shared latent space via unified representation techniques,⁷⁴ enabling cross-regional sequence correlation and automated chronostratigraphic framework construction. For lithology and mineral identification, fine-grained image recognition and multi-channel feature extraction capture subtle rock properties from logging data for automated classification,⁷⁵ while integrating optical and spectral data to invert mineral components. For underground 3D geological modeling, generative architectures leverage geological priors to generate plausible high-resolution 3D models from sparse drillhole and seismic data,⁷⁶ converting interpolation and inversion into efficient conditional generation.

For mineralization prediction, knowledge graphs and large language models mine implicit correlations in exploration data, tectonic settings and mineralization theories to optimize prospecting targets, advancing from data-driven to knowledge discovery approaches.⁷⁷

Some representative technical pathways for geological foundation models address the demands of geological knowledge enhancement and scenario-driven tasks. For instance, GeoGalactica,⁵⁵ a 30-billion-parameter language model, was pre-trained on a massive 65-billion-token geoscience corpus and instruction-tuned with professional datasets to facilitate geological knowledge synthesis and reasoning. Despite breakthroughs in data fusion and automation, large-scale deployment of geological models faces three key bottlenecks. First, unstructured knowledge processing remains challenging. Spatial-temporal relationships and symbolic logic in geological maps and stratigraphic sections are hard to structure,⁷⁸ and the lack of multi-scale knowledge graph-model integration impairs causal reasoning for complex geological processes.⁷⁹ Second, 3D modeling faces efficiency-fidelity trade-offs. Heterogeneous subsurface bodies require multi-dimensional, accurate modeling, yet no universal metrics exist to verify whether generated models meet geological constraints like tectonic continuity.⁸⁰ Third, the black-box nature⁸¹ and poor physical compliance of models hinder integration with sedimentological and tectonic theories. While physics-informed learning provides solutions, the scalability, hyperparameter tuning and computational costs remain problematic for nonlinear geological systems.

Geophysical foundation models

In contrast to the unstructured data of geology, geophysics relies on massive, high-dimensional waveform and field data. Geophysical foundation models aim to address challenges related to the ill-posed nature of geophysical inversion, slow inversion, and incomplete field reconstruction. Their tasks include seismic wave utilization and underground structure recognition, e.g., seismic imaging, inversion acceleration, and field reconstruction. TransQuake⁸² and WaveNet-based models⁸³ can improve seismic data analysis and complex waveform recognition accuracy, laying the foundation for subsequent model development. DiTingMotion⁵⁸ is built on large-scale seismic waveform data⁸⁴ and achieves seismic first-motion polarity classification. SeisCLIP⁵⁹ presents a seismology foundation model pretrained via contrastive learning on multimodal seismic spectra and event metadata, achieving competitive performance in event classification, localization, and focal mechanism analysis. SeisT⁶⁰ employs a modular transformer architecture to process seismograms and delivers gains in S-phase picking and in magnitude estimation.

Despite this progress, their large-scale application still faces bottlenecks. First, computational costs remain high. Seismic wave propagation and subsurface structure inversion require processing massive, high-dimensional spatiotemporal data. As model parameters and data dimensions increase, traditional computational resources struggle to meet operational demands. Second, limited interpretability makes it difficult to explain model outputs in physical terms. This complicates the interpretation of seismic imaging and inversion results and limits their use in decision-making. Third, generating deep structural models is challenging. Due to the scarcity of ultra-high-resolution data, existing models struggle to reconstruct dynamic changes in deep structures.^{85,86} Their simulation capabilities are limited, especially when data are incomplete.

Atmospheric foundation models

Benefiting from dense, homogeneously gridded reanalysis datasets and well-defined fluid dynamics, atmospheric science represents one of the earliest and most accomplished domains for AI foundation model applications. The high-dimensional, nonlinear, and chaotic nature of weather and climate prediction makes it an ideal scenario for validating AI's ability to model complex systems. With the release of large-scale datasets like the fifth-generation European Centre for Medium-Range Weather Forecasts (ECMWF) reanalysis (ERA5)⁸⁷ and Coupled Model Intercomparison Project Phase 6 (CMIP6),⁸⁸ coupled with the integration of high-performance computing, foundation models have achieved a leap from empirical substitution to physical fusion over the past five years.^{14,15} For selected variables and lead times, models such as GraphCast⁵¹ and Pangu-Weather⁵² have reported competi-

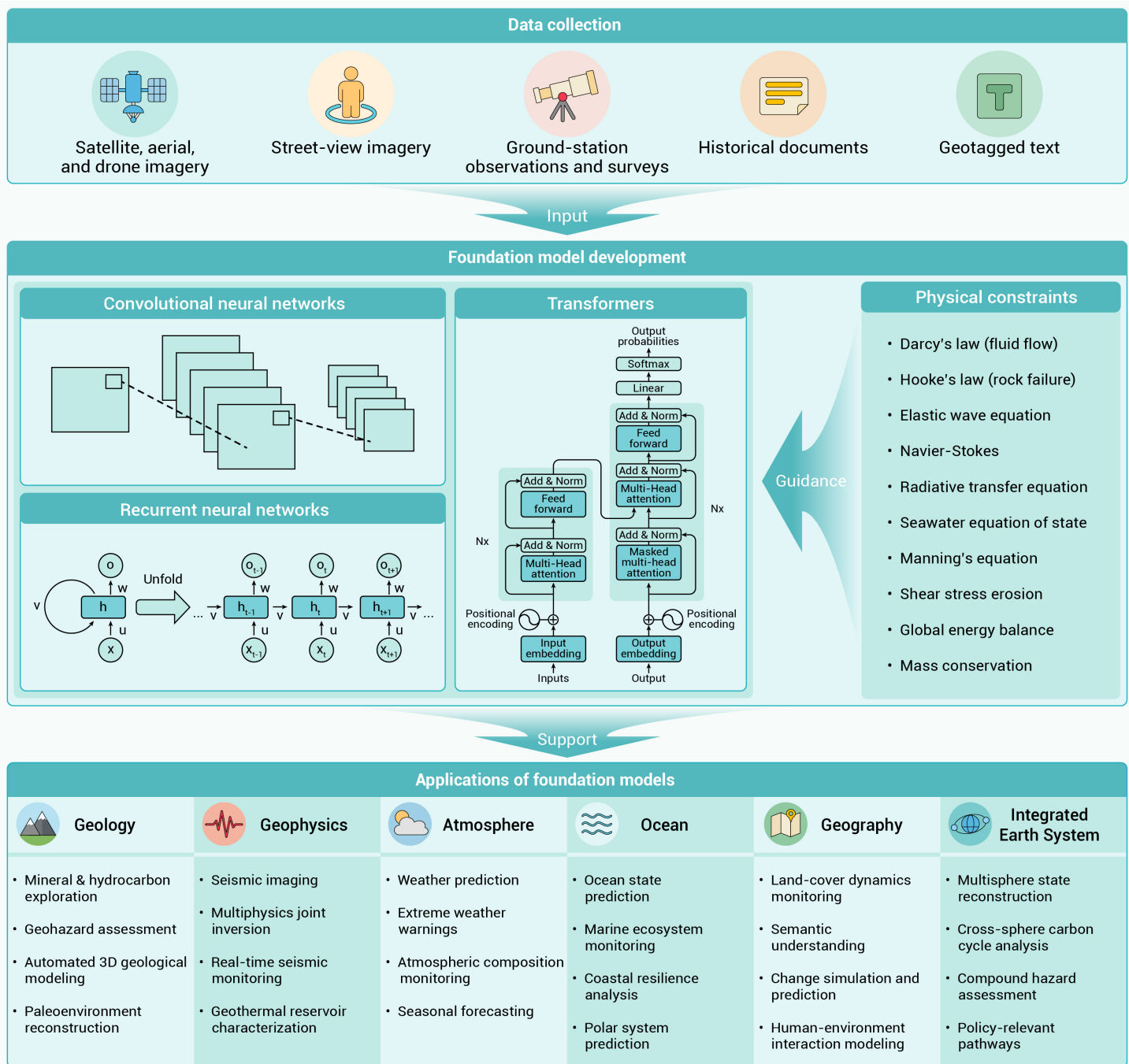


Figure 1. Applications of foundation models in ESS

tive forecast accuracy and fast inference relative to the numerical baselines used in their studies.^{51,52} Probabilistic forecasting and extreme-event applications provide further opportunities, although performance depends on the event, metric, and evaluation setting.^{52,89}

Atmospheric AI models have been widely developed for medium-range forecasting and extreme-event prediction. Operational systems such as ECMWF's Integrated Forecasting System (IFS) and the China Meteorological Administration's Global/Regional Assimilation and Prediction System (GRAPES) combine data assimilation, physical dynamics, uncertainty estimation, and maintained production workflows. AI models can approximate selected forecast mappings from reanalysis at low inference cost, but they usually depend on initial states produced by assimilation and reanalysis.¹⁶ GraphCast⁵¹ reported skill relative to the operational IFS used in its study across many evaluated variables and lead times up to 10 days, with inference completed in under one minute on the reported hardware. This timing covers forecast-model inference, not training, observation processing, data

assimilation, or the full operational workflow. Results reported for FourCast-Net⁵³ and Pangu-Weather⁵² are likewise conditional on the variables, resolution, lead time, metric, and reference configuration used in each study.⁹⁰

The generalized application of atmospheric foundation models still faces multiple bottlenecks. First, there is insufficient characterization of mesoscale and small-scale processes. Training data typically have spatial resolutions between 0.25° and 0.5°, making it difficult to capture local fine-scale structures. Additionally, models exhibit unstable responses to nonlinear abrupt events.⁹¹ Second, long-term evolution capabilities are inadequate. Multi-step prediction errors accumulate rapidly, and energy or climate drift phenomena often occur during long-term integration. Third, significant computational power and energy consumption constraints exist. Training foundation models requires massive graphics processing unit resources, yet meteorological agencies lack unified computing platforms and energy consumption assessment systems, leading to fragmented resource allocation. Fourth, insufficient coupling with operational systems persists. Hybrid models demonstrate that

Table 1. Representative models in Earth system science

Field	Representative model	Model category	Task	Characteristics
Geology	GeoBERT ⁵⁴ (2022)	Domain-specific pretrained model	Knowledge preprocessing	Secondary pre-training, optimized for geological text reasoning
	GeoGalactica ⁵⁵ (2024)	Domain-specific pretrained model	Geoscience Knowledge Synthesis	Leveraging 30B-parameter domain-adaptive language modeling with a 65B-token geoscience corpus and professional instruction tuning for specialized knowledge extraction and reasoning
	GeoProspect ⁵⁶ (2025)	Domain-specific pretrained model	Knowledge understanding and question answering	High-quality geological corpus, task-driven grouped mixture of low-rank adaptation experts
	Geo-MMRAG ⁵⁷ (2025)	Task-oriented AI model	Lithology identification	Multi-modal framework injecting geological knowledge (texts, images) into pre-trained models without retraining or large labeled datasets
Geophysics	DiTingMotion ⁵⁸ (2023)	Task-oriented AI model	Seismic wave analysis and inversion	Neural network-integrated hyperparameter imaging for conventional seismic data
	SeisCLIP ⁵⁹ (2024)	Foundation model	Multimodal seismic representation learning	Contrastive learning on spectra and event metadata for efficient fine-tuning with limited labels and strong regional transferability
	SeisT ⁶⁰ (2024)	Task-oriented AI model	Multi-task earthquake monitoring	Modular transformer delivering state-of-the-art performance across seven monitoring tasks with robust out-of-distribution generalization
	Seismic foundation model ⁶¹ (2025)	Foundation model	General seismic feature extraction	Self-supervised pretraining on over 2.2 million unlabeled images for universal cross-task representations
Atmosphere	Pangu-Weather ⁶² (2022)	Task-oriented AI model	Global End-to-End Forecasting	Based on Swin Transformer, supporting multi-scale spatio-temporal dependency learning and rapid inference
	GenCast ⁶² (2025)	Task-oriented AI model	Probabilistic forecasting	Generating multi-member ensemble forecasts based on the diffusion-based generation framework to improve the probability characterization of extreme events
	FuXi-Weather ⁶³ (2025)	Task-oriented AI model	Short-term and Long-term Forecasting	Optimizing Stability and Reliability Through Energy Conservation Regularization and Geophysical Guidance Modules
	CorrDiff ⁶⁴ (2025)	Task-oriented AI model	Kilometer-scale atmospheric downscaling	Physically consistent high-resolution weather ensembles through two-stage residual diffusion modeling that recovers multi-scale power law relationships
Ocean	AI-GOMS ⁶⁵ (2023)	Domain-specific pretrained model	Multi-task global ocean modeling	Backbone-plus-modules framework; Fourier masked autoencoder backbone; supports downscaling, waves, and biogeochemical coupling
	ORCA-DL ⁶⁶ (2025)	Task-oriented AI model	Global 3D ocean state forecasting	Transformer global forecaster; separates predictable signals from forcing; improves long-horizon stability for multivariate states
	FuXi-Ocean ⁶⁷ (2025)	Task-oriented AI model	Eddy-resolving short-term global ocean forecasting	Eddy-resolving sub-daily forecasting; mitigates autoregressive error accumulation via multi-timescale design; fast inference
	LangYa ^{68,69} (2026)	Task-oriented AI model	Global multi-variable ocean forecasting	Large language model-based lead-time embedding and asynchronous atmospheric-forcing integration; ocean self-attention and adaptive loss improve forecast stability and thermocline representation
Geography	SkySense ^{40,70} (2024, 2025)	Foundation model	Multi-modal remote sensing image interpretation	Billion-scale foundation model pre-trained on large-scale optical-SAR temporal sequences using factorized spatiotemporal encoding
	TerraMind ⁴¹ (2025)	Foundation model	Multimodal Generation and Prediction	Generative multimodal deep learning model for Earth Observation; excels in various tasks, supports "Thinking-in-Modalities"
	AlphaEarth ⁴² (2025)	Foundation model	Universal feature representation	Compressing multi-modal and multi-temporal Earth observation data into 64-dim features for each point on the global land surface and coastal areas
	Prithvi-EO-2.0 ⁷¹ (2026)	Foundation model	Earth observation applications	Pretrained on 4.2 million global time-series samples from NASA's Harmonized Landsat and Sentinel-2 imagery, with applications in disaster response, crop mapping, and ecosystem monitoring
Integrated Earth System	Aurora ¹² (2025)	Foundation model	Cross-sphere climate modeling, mid-range weather forecasting	Pretrained on heterogeneous geophysical data with adaptation to weather, air-quality, ocean-wave, and tropical-cyclone forecasting
	ESFM ⁷² (2026)	Foundation model	Forecasting with heterogeneous and partially observed Earth-system data	Extends an Aurora-based backbone to integrate gridded, satellite, and station data, supporting missing inputs and downstream adaptation; current demonstrations primarily concern atmospheric applications
	NIVA ⁷³ (2026)	Foundation model	Ocean-atmosphere representation learning and climate-index prediction	An emerging multimodal framework trained on Earth-system simulations; current proof-of-concept validation covers ocean-atmosphere interactions rather than the full atmosphere-ocean-ice-land system

learned components can be coupled with numerical dynamics,⁹² but their operational use still requires observational and simulation-based validation of stability, traceability, and robustness under extreme conditions.^{16,90} Future efforts should focus on multi-scale physical embedding, self-supervised training, cross-layer modeling,⁹³ and building explainability mechanisms to drive meteorological foundation models toward truly reshaping how we understand the Earth system. Moreover, such foundation models hold potential for applications beyond traditional weather forecasting, including air quality prediction, disaster response, and broader climate science research.

Ocean foundation models

Ocean science has witnessed a transformative integration of foundation models into ocean simulation and forecast systems, reflecting a clear trend toward cross-scale and multidisciplinary coupling. Rather than focusing on isolated tasks, current research is evolving toward a systematic framework encompassing diverse marine domains. These primary applications can be categorized into key ocean variable forecasting (such as temperature, salinity, and currents), multi-scale ocean phenomenon prediction (e.g., mesoscale eddies and internal waves), ocean disaster early warning (including storm surges and extreme waves), long-term climate simulation and change projection, and marine ecological monitoring. This hierarchical progression underscores the potential of foundation models to capture the complex and interconnected dynamics of the global ocean system.

Ocean foundation models focus on fast prediction of key ocean variables and high-resolution extrapolation, and the landscape is increasingly converging from isolated task models toward a few complementary paradigms. For example, Ocean Reliable foreCAST via Deep Learning (ORCA-DL)⁶⁶ introduces a data-driven 3-D global ocean model for seasonal-to-decadal prediction and reports robust representation of multivariate ocean dynamics and extremes, together with stable long-horizon rollouts. FuXi-Ocean⁶⁷ targets six-hourly forecasts at eddy-resolving 1/12° resolution and uses a Mixture-of-Time mechanism to reduce cumulative errors in sequential prediction.

Alongside end-to-end forecasting, a system-oriented line of work increasingly adopts the combination of a foundation backbone and lightweight downstream modules to improve reuse and scalability. Large AI-Driven Global Ocean Modeling System (AI-GOMS)⁶⁵ uses a Fourier masked autoencoder as the backbone for multi-depth ocean-variable prediction, and then extends task coverage via downstream modules for regional downscaling, wave decoding, and biogeochemical coupling. For regional digital twins and fast simulation, OceanNet⁹⁴ builds on a Fourier neural operator and a predictor-evaluate-corrector integration scheme to mitigate autoregressive error growth and improve long-horizon stability. For assimilation/interpolation-integrated modeling, 4DVarNet-SSH⁹⁵ parameterizes the update process of variational data assimilation as a trainable network, achieving end-to-end spatiotemporal interpolation of satellite altimetry observations. Language-based capabilities are also being explored. For instance, OceanRep⁹⁶ learns ocean-dynamics representations using space-time tokens and vision-transformer variants, while OceanGPT⁹⁷ introduces OceanBench and evaluates ocean-science question answering and task reasoning to support language-side applications in the data-knowledge-decision workflow.

Key challenges include data heterogeneity, physical constraints, and deployment. Sampling strategies, spatial resolutions, and error characteristics differ substantially in ocean observations, which can easily lead to noise propagation and data inconsistencies.⁹⁸ Meanwhile, observations are imbalanced in both space and depth. Coastal/nearshore and upper-ocean regions are relatively well observed, while the observations in the deep and open ocean remain sparse.⁹⁹ In long-horizon rolling prediction, errors are more likely to accumulate with lead time and manifest as systematic drift, due to limited incorporation or diagnostic enforcement of fundamental dynamics and conservation laws. Therefore, how to incorporate physical priors in a learnable manner remains a key question. 4DVarNet-SSH⁹⁵ suggests that formulating variational data assimilation as a trainable structure can help to integrate observational constraints and prior information within a unified system. Furthermore, in operational ocean services, models often run under strict production windows and constrained computational budgets. High-resolution forecasting and multi-source data processing can significantly increase memory use and latency, while long rollouts can amplify drift.¹⁰⁰

Geographic foundation models

Geographic foundation models discussed here primarily comprise geospatial and Earth observation models that leverage multi-source remote sensing imagery, aiming to integrate large-scale spatiotemporal data and geographic knowledge with AI. They are grouped under geography because they can support the representation and analysis of Earth-surface features and spatial patterns. At the perception level, these models attempt to recognize and understand various static and dynamic surface features, including buildings, roads, water bodies, land use, and forests, and can perform tasks like change detection, semantic segmentation, and 3D reconstruction. At the semantic understanding level, further development aims to enable these models to integrate geographic knowledge graphs, infer spatial relationships, and interpret geographic text and maps. At the predictive analytics level, potential applications include urban expansion forecasting, climate change simulation, and ecosystem evolution assessment, requiring spatiotemporal modeling capabilities. At the decision support level, these models could support decision-making for land use planning, disaster emergency response, and resource allocation, provided that their outputs are validated and combined with relevant physical and socioeconomic information.

Representative models illustrate advances in multimodal integration and reusable Earth-surface representations. For instance, models in the Prithvi series^{71,101} demonstrate strong transfer learning capabilities by fusing multi-temporal Landsat and Sentinel-2 data at 30-m spatial resolution. TerraMind⁴¹ further achieves the fusion of nine distinct types of Earth observation data, enabling generation of arbitrary target modalities from diverse input modalities. Models in the SkySense series^{40,70} are multi-modal remote sensing foundation models with strong performance in land cover classification and change detection tasks. The AlphaEarth foundation model⁴² integrates multi-source Earth observation data and texts by constructing a unified spatiotemporal embedding space, achieving strong generalization performance with limited labels. Similar to AlphaEarth, multiple foundation models, e.g., Tessera,¹⁰² OlmoEarth,¹⁰³ and the Embedded Seamless Data (ESD) model,¹⁰⁴ have been developed to provide unified representations of the Earth's surface at different spatial resolutions.

These models have made significant progress but still face limitations, e.g., weak humanistic semantic modeling and insufficient dynamic evolution capabilities. Current models typically integrate humanistic semantics through simple statistical processing, lacking sufficient semantic depth and contextual awareness, resulting in inaccuracies when they encounter complex socioeconomic contexts and humanistic environments. Furthermore, most models assume a relatively stable geographic environment, overlooking the nonlinear and time-varying nature of geographic processes, hindering their ability to describe and predict dynamic evolutionary processes.

Integrated Earth system foundation models

Integrated Earth system foundation models are evolving from domain-oriented atmospheric and climate models toward more general representations of the Earth system. Early foundation models such as ClimaX¹¹ and AtmoRep¹⁰⁵ demonstrated the potential of large-scale pre-training to learn transferable representations from heterogeneous atmospheric and climate data, enabling adaptation to multiple downstream tasks such as forecasting, downscaling, interpolation, and model correction. More recent models have expanded both the scale of pre-training and the range of supported applications. Prithvi WxC¹⁰⁶ learns general representations from multivariate weather and climate data and can be adapted to forecasting, downscaling, and physical parameterization tasks, while WeatherGFM¹⁰⁷ adopts a generalist framework to address different weather-related tasks within a unified model.

Aurora¹² further broadens this paradigm by pre-training on large-scale heterogeneous geophysical datasets and adapting the learned representations to weather forecasting, air-quality prediction, tropical cyclone tracking, and ocean-wave forecasting. Aurora¹² reports substantial inference-time acceleration relative to the numerical baselines used there, but the result depends on the task, resolution, hardware, comparator, and components included in the timing. Recent studies explore complementary directions toward broader Earth system representations. For instance, Earth System Foundation Model (ESFM)⁷² extends an Aurora-based architecture to integrate heterogeneous gridded, satellite, and station data, including inputs with

missing observations. Its demonstrated capabilities emphasize flexible data integration and forecasting, with evaluations primarily addressing atmospheric applications rather than fully coupled Earth system dynamics. NIVA⁷³ more directly investigates representations of interacting Earth system components. Although its proposed framework targets atmosphere, ocean, ice, and land, the reported proof of concept considers ocean-atmosphere coupling and evaluates the prediction of major climate indices. Their capabilities should be distinguished from those of complementary modeling approaches. For instance, the ultra-high-resolution ICOSahedral Non-hydrostatic (ICON) model¹⁰⁸ is an Earth system model that couples atmosphere, land, ocean, sea ice, and biogeochemistry and is included as a physics-based reference model. TritonCast¹⁰⁹ is discussed as a task-oriented AI approach to long-horizon forecasting. As a complementary process-based perspective, studies of the aviation and aerospace transition zone highlight coupled dynamical and chemical processes and observational gaps that constrain modeling.⁹³ These challenges motivate cross-layer process representation and validation, without establishing that foundation models have already resolved them.

Despite progress in transferable atmospheric representations, extending foundation models to fully coupled Earth system simulation remains an open challenge. Experience with conventional Earth system models illustrates the practical demands of coupling independently developed atmosphere, ocean, land, and biogeochemical components.¹¹⁰ Differences in spatial grids, variable definitions, and temporal resolution complicate integration and require careful treatment. These software and discretization issues should be distinguished from the scientific challenge of representing interactions and feedbacks across spheres. For foundation models, a key research objective is to develop transferable representations that can support such interactions while preserving physical consistency. Evaluation should therefore address both component performance and coupled system behavior, drawing on established Earth system model evaluation frameworks.

Foundation models can provide fast inference, reusable representations, and efficient tools for emulation, downscaling, bias correction, parameterization, and ensemble generation. However, their performance depends strongly on the training data and may deteriorate under distribution shifts, extreme events, or long integrations. Numerical weather prediction and coupled Earth system models remain necessary for physical process representation, scenario experiments, and operational forecasting. Data assimilation and reanalysis provide physically consistent state estimates and training data,^{87,98} while process-based models support mechanistic interpretation. Hybrid physics-AI approaches combine learned components with numerical solvers or physical constraints,⁹² although their stability and coupling remain challenging. Foundation models should therefore be treated as complements to established infrastructure unless replacement is demonstrated for the same task, inputs, benchmark, forecast horizon, and computational boundary. Table 2 summarizes these differences.

Transfer potential is qualitatively assessed from the studies reviewed in Section 3, considering demonstrated reuse across tasks, regions, datasets, and resolutions. “High” and “Moderate” indicate broader and more restricted reuse within evaluated settings, respectively; other descriptions identify task, regional, or cross-sphere evidence limitations. These assessments are not rankings from a common quantitative benchmark.

COMMON CHALLENGES

The high complexity and multi-scale coupling characteristics of Earth System Science make it a significant challenge to build foundation models with scientific credibility and reasoning capability. Current research remains constrained by five challenges, i.e., AI-ready Earth data dilemma, representation space fragmentation, physical inconsistency, scientific validation deficiency, and ecosystem barriers.

AI-ready Earth data dilemma

The essence of data challenges lies in the contradiction between the complexity of the Earth system and existing data structures. Observations and records of the Earth system exhibit highly multi-source, heterogeneous, and multi-scale characteristics. Spatio-temporal imbalances in data coverage further amplify model uncertainty. Temporally, the evolution of observa-

tional instruments and measurement standards introduces changes in temporal data consistency, complicating the interpretation of long-term variability.⁸⁷ Large-scale Earth system data are often non-stationary and exhibit strong regional biases, which can cause training to converge to local optima and reduce model extrapolation capabilities under shifting climate baselines. Simultaneously, vast amounts of latent geoscience knowledge remain in unstructured literature, maps, and tables, lacking unified semantic descriptions and metadata standards,¹¹¹ thereby making data alignment challenging. More fundamentally, Earth system data exhibit significant uncertainties and biases. Accumulated data noise arises from observational errors, instrument drift, and variations in processing chains. When multi-source data are fused, error propagation and nonlinear interactions may distort relationships between physical quantities. Data accessibility and resource distribution are also uneven, and differences in data availability and research infrastructure can constrain reuse and participation across institutions and regions.¹¹²

Representation space fragmentation

The Earth system spans processes ranging from microscopic turbulence to planetary-scale dynamics, forming temporal-spatial gradients that no single characterization framework can fully capture. Extending domain-oriented representations to integrated Earth system modeling requires alignment across variables, scales, and physical processes.¹³ Scale mismatches further hinder both inference and generation. Local statistics rarely generalize to global behavior, while global models often smooth out fine-scale variability. Traditional hierarchical approaches rely on empirical parameterizations to approximate multiscale feedbacks, such as turbulence influencing circulation or surface fluxes affecting climate equilibrium, and their formulation remains a source of modeling challenges.^{37,38} Resolution disparities between observations and models also distort feature extraction and physical signal representation, and multimodal datasets magnify heterogeneity when shared metadata and semantic standards are lacking.¹¹¹ At a conceptual level, Earth system representation requires rethinking underlying cognitive frameworks.¹¹³ Nonlinear geophysical processes imply that local variations do not necessarily aggregate into global patterns, and global means do not necessarily reveal local mechanisms. Consequently, coupling methods and learned representations should be evaluated together rather than assuming that accurate component models guarantee accurate system behavior.^{13,110} Achieving scale continuity, semantic consistency, and dynamical coordination thus becomes a core requirement for advancing toward a unified Earth science system.¹¹⁰

Physical inconsistency

Ensuring physical consistency remains a key challenge for Earth System Science foundation models. Traditional numerical models enforce conservation laws and dynamic constraints through explicit equations, whereas neural networks mainly optimize statistical objectives. As a result, they may produce physically inconsistent outputs when operating outside the training distribution, especially under extreme or sparsely observed conditions. In the highly nonlinear Earth system, small deviations in atmospheric, ocean, or land-surface processes can propagate across spheres, amplifying errors along coupled pathways. Violations of mass, energy, or momentum conservation often lead to long-term integration drift, undermining interpretability and limiting model credibility as scientific tools. Recent developments in physics-informed machine learning have introduced mechanisms to mitigate these issues. Analytic constraints can enforce selected physical relationships in neural networks,^{47,114,115} while physics-informed methods and hybrid dynamical models offer complementary approaches to incorporating physical information.^{92,116} Studies of geophysical inverse problems and conservation-aware learning provide task-specific evidence for the value of physical priors and constraints.¹¹⁷ Nevertheless, scaling such methods to global Earth system complexity remains difficult due to computational cost, multiple interacting conservation laws, and the stiffness of coupled physical processes.

Scientific validation deficiency

Model evaluation is fundamental to scientific modeling,^{118,119} yet current validation frameworks for foundation models insufficiently capture scientific reliability. A core limitation is the lack of criteria that can jointly assess predic-

Table 2. Cross-domain factors shaping model capability and scientific credibility

Domain	Dominant data regime	Continuity/Coverage	Transfer potential	Physical constraints	Validation practice	Principal limitation
Geology	Images, cores, logs, maps, text	Sparse; heterogeneous; often proprietary	Mostly regional/task dependent	Stratigraphy, topology, constitutive relations	Expert interpretation and local datasets	Limited standardized benchmarks and cross-region transfer
Geophysics	Waveforms and field measurements	Large volume but station/region dependent	Moderate within instrument/task families	Wave propagation and inverse-problem physics	Event catalogues, synthetic tests, field cases	Ill-posed inversion and distribution shift
Atmosphere	Global reanalysis and observations	Dense, gridded, multidecadal	High within benchmarked variables/resolution	Fluid dynamics and conservation diagnostics	Operational reanalysis benchmarks	Dependence on assimilated initial states and extreme-event robustness
Ocean	Profiles, satellites, reanalysis, simulations	Surface rich; deep ocean sparse	Moderate; resolution and basin dependent	Mass, energy, tracer conservation	Observations, reanalysis, process diagnostics	Depth sparsity and long-horizon drift
Geography	Multimodal Earth observation and geospatial data	Broad but geographically and semantically uneven	High for perception; weaker for human–environment processes	Geometry, topology, spatial context	Downstream remote-sensing benchmarks	Regional bias and limited socioeconomic semantics
Integrated Earth System	Coupled simulations and multi-sphere observation	Heterogeneous across components	Currently limited across spheres	Coupled conservation and feedbacks	Earth system model intercomparison, observations, process budgets	Representation mismatch and physically consistent coupling

tive accuracy and mechanistic plausibility. Traditional evaluations rely on statistical error metrics, assuming numerical agreement indicates validity. However, for complex Earth systems, models may reproduce observed patterns while misrepresenting underlying causal processes or feedback mechanisms, resulting in outcomes that are numerically correct but scientifically flawed.¹¹² Validation is further complicated by nonlinear, multiscale Earth system processes. Evaluation therefore requires process-oriented diagnostics alongside aggregate prediction scores, informed by established Earth system model evaluation frameworks.^{118,119} Meanwhile, the stochasticity of AI training will introduce reproducibility challenges. For instance, variations in initialization, data selection, and optimization procedures can yield divergent outcomes under nominally identical conditions.¹¹⁴ Differences across research groups in preprocessing and evaluation metrics will weaken comparability and hinder the formation of unified benchmarks.^{90,118} Uncertainty and limited interpretability also undermine scientific credibility. Structural assumptions and parameter choices contribute to uncertainty that is difficult to isolate in high-dimensional models.¹¹⁵ Moreover, the absence of transparent causal validation^{120,121} may limit the scientific falsifiability of foundation models in ESS.

Ecosystem barriers

Earth system research remains highly fragmented in resource allocation, data sharing, and cross-institutional collaboration, creating structural barriers that impede the advancement of model capabilities. Data silos, inconsistent semantic and metadata standards, and the concentration of high-quality observational datasets and advanced computing facilities in a handful of countries and institutions hinder the development of reusable, large-scale foundation models, while compromising the overall efficiency and equity of global research endeavors.¹¹² Computational demands create sustainability challenges for both foundation-model development and weather and climate computing.¹²² Carbon estimates depend on the model, electricity supply, and accounting boundary. For example, the final training of the BLOOM language model was estimated to emit 24.7 tonnes of CO₂ equivalents from dynamic power consumption alone, increasing to 50.5 tonnes when embodied equipment emissions and idle power consumption were also included.¹²³ These values illustrate the importance of transparent accounting rather than providing an estimate for Earth system foundation models. Institutional tensions between model openness and intellectual property protection, coupled with

the lack of unified data standards and interoperable interface protocols, further obstruct cross-institutional and transnational data exchange and model integration.^{111,112} Earth system science has long been compartmentalized along disciplinary boundaries, while the development of foundation models demands deep interdisciplinary coupling across geophysics, climate science, geography, information science, and even social science domains. When cross-domain collaboration entails prohibitive costs and technical standards remain unaligned, even genuine technological breakthroughs struggle to catalyze transformative paradigm shifts in Earth system research.

FUTURE PATHS

The six future paths respond to the five challenges in different but overlapping ways. A unified Earth representation engine addresses the heterogeneity of Earth observation data and the fragmentation of representation spaces across domains and scales. Physics-embedded neural architectures respond directly to physical inconsistency by incorporating conservation laws, governing equations, geometric symmetries, or other physical constraints into model training and inference.⁹² Earth system knowledge graphs can improve semantic interoperability, evidence traceability, and model auditability. Multi-scale generative inversion addresses sparse observations and scale mismatches by reconstructing fine-scale states from incomplete or coarse information.¹²⁴ Causal reasoning can support scientific validation by examining hypothesized relationships under explicit assumptions about confounding, interventions, and environmental change. Human-AI collaborative interfaces are proposed to support reproducibility, expert oversight, evidence checking, and audit trails, drawing on reproducible scientific workflows and experience with automated laboratories.¹²⁵ These paths are complementary, as a single challenge may require several responses, while one research path may simultaneously address limitations in data, representation, validation, and scientific practice.

Near-term priorities include harmonized data and metadata standards, transparent benchmarks, physical consistency diagnostics, calibrated uncertainty estimates, and reproducible hybrid workflows.^{111,90} These measures can be implemented using current observations, numerical models, and evaluation systems. Cross-sphere causal models globally coupled generative digital twins, autonomous scientific discovery, and planetary-scale governance remain longer-term goals. Their feasibility has not yet been demonstrated across the full Earth system. Progress toward these goals will require out-of-

distribution testing, conservation checks, uncertainty calibration, comparison with operational systems, and prospective evaluation in real scientific workflows. Human-AI collaborative discovery should likewise be introduced first in auditable tasks such as evidence retrieval, model diagnosis, and hypothesis screening before it is extended to autonomous experimentation or decision support.

Unified Earth representation engine

A unified Earth representation engine aims to achieve reversible alignment between images, numerical fields, and text through a shared latent semantic space, providing a semantic foundation for cross-layer transfer and multi-task generalization. Achieving this goal requires synergistic innovation at both data and algorithmic levels. At the data level, improving access to interoperable geoscience data resources is a priority, along with unified metadata standards and cross-platform interface protocols. At the algorithmic level, structured priors such as geographic location, time series, and layer attributes should be fused during pre-training to construct a multimodal embedding space. Concurrently, integrating meta-learning techniques¹²⁶ can further enhance the model's generalization capabilities across diverse tasks. Notably, Prithvi-EO-2.0⁷¹ demonstrates exceptional transfer learning capabilities across multi-source remote sensing tasks through unsupervised pre-training. AlphaEarth⁴² innovatively encodes multi-source Earth observation data into 64-dimensional embedding vectors, forming a globally consistent geospatial fingerprint, which significantly reduces computational and annotation costs. In the future, unified representation engines should further expand their 3D and 4D voxel representation capabilities to enhance understanding of cross-scale processes such as tectonic evolution and climate change. Simultaneously, they should incorporate physical constraints and uncertainty quantification mechanisms to ensure scientific credibility and interpretability.

Physics-embedded neural architecture

Physical constraints can be introduced across optimization, structure, and generation levels. At the optimization level, physics-informed neural networks¹¹⁶ incorporate partial differential equation residuals, boundary conditions, and conservation laws as regularization terms within the loss function, encouraging agreement with the specified physical equations. At the structural level, E(3)-equivariant neural networks encode geometric symmetries, as demonstrated for interatomic potentials.¹²⁷ Their use in Earth system applications requires task-specific design and validation. At the generation level, physics-informed diffusion methods have been explored for flow-field reconstruction,¹²⁸ although reconstruction performance alone does not establish conservation or stability during long integrations. Furthermore, hybrid modeling strategies combining AI and numerical models are a significant trend. AI can approximate unresolved processes or correct model components, while a differentiable dynamical solver retains an explicit representation of resolved physical processes, enabling joint optimization as demonstrated by NeuralGCM.⁹² Different forms of physical constraints can be imposed with varying degrees of strictness depending on data availability and modeling objectives. Future efforts should establish a training strategy that combines soft and hard constraints, by employing softened constraints in data-dense regions to enhance model expressiveness and strengthening hard constraints in observation-sparse areas to stabilize extrapolation performance. Concurrently, integrating physical consistency scores into model evaluation frameworks will embed credibility as an intrinsic property, driving the shift from statistical fitting to mechanistic understanding.

Earth system knowledge graph (GeoKG)

Centuries of accumulated qualitative knowledge and theoretical frameworks in Earth System Science remain fragmented across literature and expert experience, inaccessible to data-driven models. Constructing the GeoKG can translate this knowledge into computable semantic networks that provide scientific priors. This requires coordinated advances in ontology modeling, knowledge extraction, and dynamic updating. In ontology modeling, mature semantic frameworks such as SWEET¹²⁹ and GeoSciML¹³⁰ can provide a conceptual framework to improve semantic consistency and interoperability across interdisciplinary data. During knowledge extraction, natural language processing and structured extraction techniques should be

combined to extract event-mechanism-evidence triplets from scientific literature, field reports, and observational data. Each knowledge unit should be annotated with spatio-temporal attributes and source information to build a high-quality, traceable knowledge base. Dynamic update mechanisms should be incorporated into the GeoKG to continually refine and expand its content as new observational data and scientific insights become available. Recent research has systematically summarized construction methods and application pathways for geoscience knowledge graphs, providing framework guidance for engineering implementation.¹³¹⁻¹³⁴ Future deep integration of GeoKG and Earth system foundation models will build a data-knowledge-decision system, supporting cross-disciplinary causal analysis and scenario simulation, and ultimately empowering models with genuine scientific discovery capabilities.

Multi-scale generative inversion

Earth system observations exhibit characteristic sparsity and multi-scale features ranging from local to planetary scales. Generative artificial intelligence offers a novel technical pathway to reconstruct high-resolution Earth system state fields from sparse observations, while aiming to preserve physical consistency in terms of mass, energy, and momentum conservation. Regarding generative models, diffusion models and autoregressive transformers can perform multiscale multivariate state reconstruction in a unified representation space through progressive denoising. Denoising diffusion provides a general generative modeling framework,¹³⁵ with domain-specific applications including atmospheric downscaling and digital-rock reconstruction. Score-based data assimilation offers a related approach for inferring state trajectories from noisy or incomplete observations.¹²⁴ To incorporate observational information, trainable variational assimilation and observation-guided generative inference provide complementary approaches.¹²⁴ Where suitable formulations are available, explicit constraints on quantities such as energy and mass can complement observational conditioning.¹¹⁷ Moving forward, multiscale generation and inversion could support the joint generation of multiple Earth system variables and integration across disciplines, contributing to the development of a digital twin of the Earth.

Causal reasoning and scenario simulation

Earth System Science foundation models should transcend statistical correlations by encoding forcing-response-feedback dynamics into structural causal models, enabling causal understanding, counterfactual scenario simulation, and quantifiable causal inference within an intervenable semantic framework. Achieving this goal requires multi-level integration. At the representation layer, structural causal models should depict direct relationships between key variables to clarify the causal structure of Earth system processes. At the learning layer, representation-learning objectives should be paired with explicit causal assumptions and, where possible, evidence from interventions or environmental variation, because statistical objectives alone are insufficient to establish causal relationships.^{120,136} At the generation layer, counterfactual generation and intervention computation should form multiple verifiable scenario paths, with physical consistency regularization preventing spurious counterfactuals. The evaluation layer should output average treatment effects, conditional effects, and their uncertainties, supplemented by sensitivity and robustness tests.¹²⁰ Causal chains should be interconnected with GeoKG, enabling each conclusion to possess traceable literature support and data evidence. Causal representation learning investigates how causally meaningful variables may be learned under additional assumptions or supervision.¹³⁶ Causal inference algorithms for complex systems demonstrate structural recognition capabilities for circulation-precipitation-extreme event chains in Earth science time-series problems.¹³⁷ Data fusion theory can provide a consistent inference framework for integrating multi-source observations and model outputs.¹³⁸ In the future, this module is expected to integrate with unified representation, physical embedding, and knowledge graphs to support the visual assessment and dynamic optimization of policy scenario libraries.

Human-AI collaborative discovery interface

Scientists should move beyond model use to guide problem formulation, critically assess AI-generated hypotheses, and validate findings. This requires

building a next-generation Human-AI collaborative discovery interface to support the entire scientific lifecycle from problem formulation to evidence retrieval. The implementation of this interface should consider data, model, interaction, execution, and governance. The data layer integrates multi-source Earth observation, experimental, and literature data, providing a unified access interface. The model layer executes complex reasoning and learning tasks. The interaction layer provides multimodal visual analysis, evidence tracing, physical consistency verification, and causal explanations. The execution layer interfaces with autonomous experimental equipment and high-performance computing platforms to achieve hypothesis validation. The governance layer manages version control, audit logs, ethical compliance, and carbon footprint accounting.¹²³ AI agents¹³⁹ have recently attracted increasing attention for their ability to coordinate multiple components through task planning, tool use, and iterative feedback under scientific supervision. Methods such as Earth AI,¹⁴⁰ Earth-Agent,¹⁴¹ and OpenEarth-Agent¹⁴² illustrate the emerging use of agent-based frameworks for geospatial reasoning and task execution. This paradigm can significantly compress the hypothesis-verification cycle.¹⁴³ Techniques like Bayesian reaction optimization efficiently boost yields through small-sample experiments.¹⁴⁴ Robotic scientist systems demonstrate automated end-to-end workflows from hypothesis generation to experimental validation.¹⁴⁵ The self-driving lab concept synthesizes pathways for AI-experiment collaboration, accelerating scientific discovery.¹²⁵ Scientists can focus on creative thinking while entrusting tedious data processing, model training, and experimental execution to the system, improving research efficiency.

CONCLUSION

This review synthesizes advances in foundation models across geology, geophysics, atmospheric science, ocean science, geography, and integrated Earth system modeling. Across these fields, reusable representations and adaptation to multiple tasks offer opportunities to connect heterogeneous observations, scientific knowledge, and predictive modeling. Progress nevertheless remains uneven, reflecting differences in observational coverage, physical constraints, and validation requirements. Data heterogeneity, fragmented representations, physical inconsistency, inadequate scientific validation, and unequal access to research infrastructure continue to constrain transfer from individual applications to coupled Earth system understanding. Addressing these limitations requires coordinated development of unified Earth representations, physics-embedded architectures, knowledge graphs, multiscale generative inversion, causal reasoning, and human-AI collaboration. Near-term efforts should prioritize interoperable data, reproducible benchmarks, conservation diagnostics, calibrated uncertainty, and hybrid workflows that integrate learned capabilities with numerical modeling and data assimilation. Evaluation should extend beyond predictive accuracy to assess robustness under distribution shifts, long-term stability, and cross-sphere feedbacks. Fully coupled digital twins, causal scenario simulation, and autonomous discovery remain research goals whose feasibility requires further evidence. The scientific contribution of foundation models will ultimately depend on whether they deliver transferable, physically credible, and independently validated capabilities that strengthen our ability to observe, explain, and anticipate Earth system change.

REFERENCES

- Steffen W., Richardson K., Rockström J., et al. (2020). The emergence and evolution of Earth system science. *Nat. Rev. Earth Environ.* 1:54–63. DOI:10.1038/s43017-019-0005-6
- Kumar A., Nagar S. and Anand S. (2021). Climate change and existential threats. Singh S., Singh P., Selvasembian R., et al. (eds.). *Global Climate Change* (Elsevier), pp: 1–31. DOI: 10.1016/B978-0-12-822928-6.00005-8
- Raymond C., Horton R.M., Zscheischler J., et al. (2020). Understanding and managing connected extreme events. *Nat. Clim. Change* 10:611–621. DOI:10.1038/s41558-020-0790-4
- Cao Y., Wang H., Su F., et al. (2026). Satellite radar and AIS reveal a 97% decline in shipping traffic through the Strait of Hormuz. *The Innovation* 7:101367. DOI:10.1016/j.xinn.2026.101367
- Li X., Feng M., Ran Y., et al. (2023). Big data in Earth system science and progress towards a digital twin. *Nat. Rev. Earth Environ.* 4:319–332. DOI:10.1038/s43017-023-00409-w
- Guo H., Wang L. and Liang D. (2016). Big earth data from space: a new engine for Earth science. *Sci. Bull.* 61:505–513. DOI:10.1007/s11434-016-1041-y
- Gray J. (2009). Jim Gray on eScience: A transformed scientific method. Hey T., Tansley S., and Tolle K.M. (eds.). *The Fourth Paradigm: Data-Intensive Scientific Discovery* (Microsoft Research), pp: xvii–xxxi.
- Radford A., Narasimhan K., Salimans T., et al. (2018). Improving language understanding by generative pre-training. https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf
- Guo D., Yang D., Zhang H., et al. (2025). DeepSeek-R1 incentivizes reasoning in LLMs through reinforcement learning. *Nature* 645:633–638. DOI:10.1038/s41586-025-09422-z
- Liu Y., Zhang K., Li Y., et al. (2024). Sora: a review on background, technology, limitations, and opportunities of large vision models. *arXiv*. <https://arxiv.org/abs/2402.17177>
- Nguyen T., Brandstetter J., Kapoor A., et al. (2023). ClimaX: A foundation model for weather and climate. *arXiv*. <https://arxiv.org/abs/2301.10343>
- Bodnar C., Bruinsma W. P., Lucic A., et al. (2025). A foundation model for the Earth system. *Nature* 641:1180–1187. DOI:10.1038/s41586-025-09005-y
- Irrgang C., Boers N., Sonnewald M., et al. (2021). Towards neural Earth system modelling by integrating artificial intelligence in Earth system science. *Nat. Mach. Intell.* 3:667–674. DOI:10.1038/s42256-021-00374-3
- Reichstein M., Camps-Valls G., Stevens B., et al. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature* 566:195–204. DOI:10.1038/s41586-019-0912-1
- Schultz M. G., Betancourt C., Gong B., et al. (2021). Can deep learning beat numerical weather prediction. *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.* 379:20200097. DOI:10.1098/rsta.2020.0097
- Dueben P. D. and Bauer P. (2018). Challenges and design choices for global weather and climate models based on machine learning. *Geosci. Model Dev.* 11:3999–4009. DOI:10.5194/gmd-11-3999-2018
- Zhang H., Xu J.J., Cui H.W., et al. (2024). When geoscience meets foundation models: Toward a general geoscience artificial intelligence system. *IEEE Geosci. Remote Sens. Mag.* 2–41. DOI:10.1109/MGRS.2024.3496478
- Karpatne A., Ebert-Uphoff I., Ravela S., et al. (2019). Machine learning for the geosciences: challenges and opportunities. *IEEE Trans. Knowl. Data Eng.* 31:1544–1554. DOI:10.1109/TKDE.2018.2861006
- Persello C., Wegner J. D., Hänsch R., et al. (2022). Deep learning and Earth observation to support the sustainable development goals: current approaches, open challenges, and future opportunities. *IEEE Geosci. Remote Sens. Mag.* 10:172–200. DOI:10.1109/MGRS.2021.3136100
- Zhao T., Wang S., Ouyang C., et al. (2024). Artificial intelligence for geoscience: progress, challenges, and perspectives. *The Innovation* 5:100691. DOI:10.1016/j.xinn.2024.100691
- Bergen K. J., Johnson P. A., de Hoop M. V., et al. (2019). Machine learning for data-driven discovery in solid Earth geoscience. *Science* 363:eaau0323. DOI:10.1126/science.aau0323
- Rhoads B.L. and Thorn C.E. (eds.) (1996). *The scientific nature of geomorphology: Proceedings of the 27th Binghamton symposium in geomorphology* (Wiley).
- Burrough P. A. (1981). Fractal dimensions of landscapes and other environmental data. *Nature* 294:240–242. DOI:10.1038/294240a0
- Chamberlin T. C. (1890). The method of multiple working hypotheses. *Science ns-15*:92–96. DOI:10.1126/science.ns-15.366.92
- Byers C. W. (1982). Stratigraphy—The fall of continuity. *J. Geol. Educ.* 30:215–221. DOI:10.5408/0022-1368-30.4.215
- Harris E. C. (1979). The laws of archaeological stratigraphy. *World Archaeol.* 11:111–117. DOI:10.1080/00438243.1979.9979753
- Short N.M. and Blair R.W. Jr. (eds.) (1986). *Geomorphology from space: A global overview of regional landforms*. Washington DC: Scientific and Technical Information Branch, National Aeronautics and Space Administration, pp: 717.
- Kondolf G.M., Piégay H., Schmitt L., et al. (2016). Geomorphic classification of rivers and streams. Kondolf G.M. and Piégay H. (eds.). *Tools in Fluvial Geomorphology* (John Wiley & Sons, Ltd), pp: 133–158. DOI: 10.1002/9781118648551.ch7
- Greene M.T. (2017). *Geology in the Nineteenth century: Changing views of a changing world*. Cornell University Press. DOI: 10.7591/9781501704741
- Bokulich A. and Oreskes N. (2017). Models in geosciences. Magnani L. and Bertolotti T. (eds.). *Springer Handbook of Model-Based Science* (Springer International Publishing), pp: 891–911. DOI: 10.1007/978-3-319-30526-4_41
- Bauer P., Thorpe A. and Brunet G. (2015). The quiet revolution of numerical weather prediction. *Nature* 525:47–55. DOI:10.1038/nature14956
- Turner A.K. and Gable C.W. (2007). A review of geological modeling. Thorleifson L.H., Berg R.C. and Russell H.A.J. (eds.). *Three-Dimensional Geologic Mapping for Groundwater Applications* (Minnesota Geological Survey), pp: 75–79.
- Johannessen O.M. (1986). Brief overview of the physical oceanography. Hurdle B.G. (ed.) *The Nordic Seas* (Springer), pp: 103–128. DOI: 10.1007/978-1-4615-8035-5_4
- Pierce D. W., Barnett T. P., Santer B. D., et al. (2009). Selecting global climate models for regional climate change studies. *Proc. Natl. Acad. Sci.* 106:8441–8446. DOI:10.1073/pnas.0900094106
- Schneider T., Lan S., Stuart A., et al. (2017). Earth system modeling 2.0: a blueprint for models that learn from observations and targeted high-resolution simulations. *Geophys. Res. Lett.* 44. DOI:10.1002/2017GL076101
- Li X., Chen Y., Liu X., et al. (2011). Concepts, methodologies, and tools of an integrated geographical simulation and optimization system. *Int. J. Geogr. Inf. Sci.*

- 25:633–655. DOI:10.1080/13658816.2010.496370
37. Fox-Kemper B., Adcroft A., Böning C. W., et al. (2019). Challenges and prospects in ocean circulation models. *Front. Mar. Sci.* **6**. DOI:10.3389/fmars.2019.00065
 38. Jakob C. (2010). Accelerating progress in global atmospheric model development through improved parameterizations: Challenges, opportunities, and strategies. *Bull. Am. Meteorol. Soc.* **91**:869–876. DOI:10.1175/2009BAMS2898.1
 39. Xu Y., Liu X., Cao X., et al. (2021). Artificial intelligence: a powerful paradigm for scientific research. *The Innovation* **2**:100179. DOI:10.1016/j.xinn.2021.100179
 40. Guo X., Lao J., Dang B., et al. (2024). SkySense: A multi-modal remote sensing foundation model towards universal interpretation for Earth observation imagery. arXiv. <https://arxiv.org/abs/2312.10115>
 41. Jakubik J., Yang F., Blumenstiel B., et al. (2025). TerraMind: Large-scale generative multimodality for Earth observation. arXiv. <https://arxiv.org/abs/2504.11171>
 42. Brown C. F., Kazmierski M. R., Pasquarella V. J., et al. (2025). AlphaEarth foundations: An embedding field model for accurate and efficient global mapping from sparse label data. arXiv. <https://arxiv.org/abs/2507.22291>
 43. Xiang Y., Jin W., Dong H., et al. (2025). ADAF: an artificial intelligence data assimilation framework for weather forecasting. *J. Adv. Model. Earth Syst.* **17**:e2024MS004839. DOI:10.1029/2024MS004839
 44. Liu Q., Chen Z. and Ma J. (2025). Foundation models for exploration geophysics. *Geophysics* **90**:H29–H52. DOI:10.1190/geo2024-0916.1
 45. Bommasani R., Hudson D. A., Adeli E., et al. (2021). On the opportunities and risks of foundation models. arXiv. <https://arxiv.org/abs/2108.07258>
 46. Plésiat É., Dunn R. J. H., Donat M. G., et al. (2024). Artificial intelligence reveals past climate extremes by reconstructing historical records. *Nat. Commun.* **15**:9191. DOI:10.1038/s41467-024-53464-2
 47. Karniadakis G. E., Kevrekidis I. G., Lu L., et al. (2021). Physics-informed machine learning. *Nat. Rev. Phys.* **3**:422–440. DOI:10.1038/s42254-021-00314-5
 48. Zhang R., Zhou L., Gao C., et al. (2024). A transformer-based coupled ocean-atmosphere model for ENSO studies. *Sci. Bull.* **69**. DOI:10.1016/j.scib.2024.04.048
 49. Zhang P., Zhou L., Li Y., et al. (2025). Self-supervised learning of invariant causal representation in heterogeneous information network. *Inf. Fusion* **123**:103246. DOI:10.1016/j.inffus.2025.103246
 50. Li X. and Guo Y. (2025). Paradigm shifts from data-intensive science to robot scientists. *Sci. Bull.* **70**:14–18. DOI:10.1016/j.scib.2024.09.029
 51. Lam R., Sanchez-Gonzalez A., Willson M., et al. (2023). Learning skillful medium-range global weather forecasting. *Science* **382**:1416–1421. DOI:10.1126/science.adf2336
 52. Bi K., Xie L., Zhang H., et al. (2023). Accurate medium-range global weather forecasting with 3D neural networks. *Nature* **619**:533–538. DOI:10.1038/s41586-023-06185-3
 53. Pathak J., Subramanian S., Harrington P., et al. (2022). FourCastNet: A global data-driven high-resolution weather model using adaptive fourier neural operators. arXiv. <https://arxiv.org/abs/2202.11214>
 54. Gao Y., Xiong Y., Wang S., et al. (2022). GeoBERT: Pre-training geospatial representation learning on point-of-interest. *Appl. Sci.* **12**:12942. DOI:10.3390/app122412942
 55. Lin Z., Deng C., Zhou L., et al. (2024). GeoGalactica: a scientific large language model in geoscience. arXiv. <https://arxiv.org/abs/2401.00434>
 56. Wu G., Wang H., Zhang K., et al. (2025). GeoProspect: A domain-specific geological large language model with enhanced continual learning. *Neurocomputing* **650**:130801. DOI:10.1016/j.neucom.2025.130801
 57. Chen X., Liu Y., Han W., et al. (2025). A vision-language foundation model-based multi-modal retrieval-augmented generation framework for remote sensing lithological recognition. *ISPRS J. Photogramm. Remote Sens.* **225**:328–340. DOI:10.1016/j.isprsjprs.2025.04.015
 58. Zhao M., Xiao Z., Zhang M., et al. (2023). DiTingMotion: a deep-learning first-motion-polarity classifier and its application to focal mechanism inversion. *Front. Earth Sci.* **11**:1103914. DOI:10.3389/feart.2023.1103914
 59. Si X., Wu X., Sheng H., et al. (2024). SeisCLIP: a seismology foundation model pre-trained by multimodal data for multipurpose seismic feature extraction. *IEEE Trans. Geosci. Remote Sens.* **62**:1–13. DOI:10.1109/TGRS.2024.3354456
 60. Li S., Yang X., Cao A., et al. (2024). SeisT: a foundational deep-learning model for earthquake monitoring tasks. *IEEE Trans. Geosci. Remote Sens.* **62**:1–15. DOI:10.1109/TGRS.2024.3371503
 61. Sheng H., Wu X., Si X., et al. (2025). Seismic foundation model: a next generation deep-learning model in geophysics. *Geophysics* **90**:IM59–IM79. DOI:10.1190/geo2024-0262.1
 62. Price I., Sanchez-Gonzalez A., Alet F., et al. (2025). Probabilistic weather forecasting with machine learning. *Nature* **637**:84–90. DOI:10.1038/s41586-024-08252-9
 63. Sun X., Zhong X., Xu X., et al. (2025). A data-to-forecast machine learning system for global weather. *Nat. Commun.* **16**:6658. DOI:10.1038/s41467-025-62024-1
 64. Mardani M., Brenowitz N., Cohen Y., et al. (2025). Residual corrective diffusion modeling for km-scale atmospheric downscaling. *Commun. Earth Environ.* **6**:124. DOI:10.1038/s43247-025-02042-5
 65. Xiong W., Xiang Y., Wu H., et al. (2023). AI-GOMS: Large AI-driven global ocean modeling system. arXiv. <https://arxiv.org/abs/2308.03152>
 66. Guo Z., Lyu P., Ling F., et al. (2025). Data-driven global ocean modeling for seasonal to decadal prediction. *Sci. Adv.* **11**:eadu2488. DOI:10.1126/sciadv.adu2488
 67. Huang Q., Niu Y., Zhong X., et al. (2026). A deep learning global ocean forecasting model with sub-daily and eddy-resolving resolution. *npj Climate and Atmospheric Science*. DOI: 10.1038/s41612-026-01444-2
 68. Yang N., Wang C., Zhao Z., et al. (2026). LangYa: a large AI model for global ocean forecasting. *Sci. Bull.* DOI:10.1016/j.scib.2026.08.001
 69. Li X., Zhang X., Ren Y., et al. (2026). Artificial intelligence for ocean phenomena forecasting. *Sci. Bull.* **71**:2632–2652. DOI:10.1016/j.scib.2026.04.002
 70. Wu K., Zhang Y., Ru L., et al. (2025). A semantic-enhanced multi-modal remote sensing foundation model for Earth observation. *Nat. Mach. Intell.* **7**:1235–1249. DOI:10.1038/s42256-025-01078-8
 71. Szwarcman D., Roy S., Fraccaro P., et al. (2026). Prithvi-EO-2.0: A versatile multimodal foundation model for Earth observation applications. *IEEE Trans. Geosci. Remote Sens.* **64**:1–20. DOI:10.1109/TGRS.2025.3642610
 72. Ozdemir F., Cheng Y., Mohebi S., et al. (2026). Earth system foundation model (ESFM): a unified framework for heterogeneous data integration and forecasting. arXiv. <https://arxiv.org/abs/2605.00850>
 73. Pal A., Sweeney A., Heyblom K., et al. (2026). NIVA: a multimodal foundation model for actionable earth system intelligence. arXiv. <https://arxiv.org/abs/2606.28546v1>
 74. Wang C., Wei X., Pan H., et al. (2024). Well logging stratigraphic correlation algorithm based on Semantic Segmentation. *Appl. Geophys.* **21**:650–666. DOI:10.1007/s11770-024-1085-8
 75. Sun Y., Pang S., Zhao Z., et al. (2024). Interpretable SHAP model combining Meta-learning and vision transformer for lithology classification using limited and unbalanced drilling data in well logging. *Nat. Resour. Res.* **33**:2545–2565. DOI:10.1007/s11053-024-10396-4
 76. Jiang C. and Lei Q. (2025). FracLab: A robust 3D DFN generator for coupled stress-flow-transport modeling of complex fractured media. *Comput. Geotech.* **186**:107444. DOI:10.1016/j.compgeo.2025.107444
 77. Yang F., Zuo R. and Kreuzer O. P. (2024). Artificial intelligence for mineral exploration: A review and perspectives on future directions from data science. *Earth-Sci. Rev.* **258**:104941. DOI:10.1016/j.earscirev.2024.104941
 78. Dagdelen J., Dunn A., Lee S., et al. (2024). Structured information extraction from scientific text with large language models. *Nat. Commun.* **15**:1418. DOI:10.1038/s41467-024-45563-x
 79. Chen Y., Tian M., Wu Q., et al. (2024). A deep learning-based method for deep information extraction from multimodal data for geological reports to support geological knowledge graph construction. *Earth Sci. Inform.* **17**:1867–1887. DOI:10.1007/s12145-023-01207-0
 80. Luo X., Sun J., Zhang R., et al. (2024). A multi-condition denoising diffusion probabilistic model controls the reconstruction of 3D digital rocks. *Comput. Geosci.* **184**:105541. DOI:10.1016/j.cageo.2024.105541
 81. McGovern A., Lagerquist R., Gagne D. J., et al. (2019). Making the black box more transparent: Understanding the physical implications of machine learning. *Bull. Am. Meteorol. Soc.* **100**:2175–2199. DOI:10.1175/BAMS-D-18-0195.1
 82. Hu Y., Zhang Q., Zhao W., et al. (2021). TransQuake: A transformer-based deep learning approach for seismic P-wave detection. *Earthq. Res. Adv.* **1**:100004. DOI:10.1016/j.eqrea.2021.100004
 83. Sun M., Yu M., Lv P., et al. (2021). Man-made threat event recognition based on distributed optical fiber vibration sensing and SE-WaveNet. *IEEE Trans. Instrum. Meas.* **70**:1–11. DOI:10.1109/TIM.2021.3081178
 84. Zhao M., Xiao Z., Chen S., et al. (2023). DiTing: a large-scale chinese seismic benchmark dataset for artificial intelligence in seismology. *Earthq. Sci.* **36**:84–94. DOI:10.1016/j.eqs.2022.01.022
 85. Ongie G., Jalal A., Metzler C. A., et al. (2020). Deep learning techniques for inverse problems in imaging. *IEEE J. Sel. Areas Inf. Theory* **1**:39–56. DOI:10.1109/JSAIT.2020.2991563
 86. Wu X., Ma J., Si X., et al. (2023). Sensing prior constraints in deep neural networks for solving exploration geophysical problems. *Proc. Natl. Acad. Sci.* **120**:e2219573120. DOI:10.1073/pnas.2219573120
 87. Hersbach H., Bell B., Berrisford P., et al. (2020). The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* **146**:1999–2049. DOI:10.1002/qj.3803
 88. Eyring V., Bony S., Meehl G. A., et al. (2016). Overview of the coupled model intercomparison project phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.* **9**:1937–1958. DOI:10.5194/gmd-9-1937-2016
 89. Camps-Valls G., Fernández-Torres M.Á., Cohrs K.H., et al. (2025). Artificial intelligence for modeling and understanding extreme weather and climate events. *Nat. Commun.* **16**:1919. DOI:10.1038/s41467-025-56573-8
 90. Rasp S., Hoyer S., Merose A., et al. (2024). WeatherBench 2: A Benchmark for the Next Generation of Data - Driven Global Weather Models. *J. Adv. Model. Earth Syst.* **16**:e2023MS004019. DOI:10.1029/2023MS004019
 91. Weyn J. A., Durran D. R. and Caruana R. (2020). Improving data - driven global weather prediction using deep convolutional neural networks on a cubed sphere. *J. Adv. Model. Earth Syst.* **12**:e2020MS002109. DOI:10.1029/2020MS002109
 92. Kochkov D., Yuval J., Langmore I., et al. (2024). Neural general circulation models for weather and climate. *Nature* **632**:1060–1066. DOI:10.1038/s41586-024-07744-y
 93. Sheng Z., He Y., Wang S., et al. (2025). Dynamics, chemistry, and modeling studies in the aviation and aerospace transition zone. *The Innovation* **6**:101012. DOI:10.1016/j.xinn.2025.101012
 94. Chattopadhyay A., Gray M., Wu T., et al. (2024). OceanNet: a principled neural operator-based digital twin for regional oceans. *Sci. Rep.* **14**:21181. DOI:10.1038/s41598-024-72145-0
 95. Beauchamp M., Febvre Q., Georgenthum H., et al. (2023). 4DVarNet-SSH: end-to-end learning of variational interpolation schemes for nadir and wide-swath satellite

- altimetry. *Geosci. Model Dev.* **16**:2119–2147. DOI:10.5194/gmd-16-2119-2023
96. Nowak K., Koldunov N., Jung T., et al. (2025). OceanRep: A foundation model for ocean dynamics. Vienna: EGU General Assembly 2025. DOI: 10.5194/egusphere-egu25-15061
 97. Bi Z., Zhang N., Xue Y., et al. (2024). OceanGPT: A large language model for ocean science tasks. Ku L.W., Martins A. and Srikanth V. (eds). Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Association for Computational Linguistics), pp: 3357–3372. DOI: 10.18653/v1/2024.acl-long.184
 98. Brajard J., Carraschi A., Bocquet M., et al. (2021). Combining data assimilation and machine learning to infer unresolved scale parametrization. *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.* **379**:20200086. DOI:10.1098/rsta.2020.0086
 99. Zilberman N. V., Thierry V., King B., et al. (2023). Observing the full ocean volume using Deep Argo floats. *Front. Mar. Sci.* **10**:1287867. DOI:10.3389/fmars.2023.1287867
 100. Wan L., Garcia Sotillo M., Bell M., et al. (2025). An introduction to operational chains in ocean forecasting. *State Planet* **5**-oprs: 15. DOI: 10.5194/sp-5-oprs-15-2025
 101. Jakubik J., Roy S., Phillips C. E., et al. (2023). Foundation Models for Generalist Geospatial Artificial Intelligence. arXiv. <https://arxiv.org/abs/2310.18660>
 102. Feng Z., Atzberger C., Jaffer S., et al. (2025). TESSERA: Temporal embeddings of surface spectra for Earth representation and analysis. arXiv. <https://arxiv.org/abs/2506.20380>
 103. Herzog H., Bastani F., Zhang Y., et al. (2026). OlmoEarth: Stable latent image modeling for multimodal Earth observation. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp: 34806–34817.
 104. Chen S., Wang J., Yuan S., et al. (2026). Democratizing planetary-scale analysis: an ultra-lightweight Earth embedding database for accurate and flexible global land monitoring. *Earth Syst. Sci. Data* **18**:5375–5398. DOI:10.5194/essd-18-5375-2026
 105. Lessig C., Luise I., Gong B., et al. (2023). AtmoRep: A stochastic model of atmosphere dynamics using large scale representation learning. arXiv. <https://arxiv.org/abs/2308.13280>
 106. Schmude J., Roy S., Trojak W., et al. (2024). Prithvi WxC: Foundation model for weather and climate. arXiv. <https://arxiv.org/abs/2409.13598>
 107. Zhao X., Zhou Z., Zhang W., et al. (2025). WeatherGFM: learning a weather generalist foundation model via In-context learning. *Int. Conf. Learn. Represent.* **2025**:80665–80685.
 108. Klocke D., Frauen C., Engels J. F., et al. (2025). Computing the full earth system at 1 km resolution. arXiv. <https://arxiv.org/abs/2511.02021>
 109. Wu H., Gao Y., Gou R., et al. (2025). Advanced long-term earth system forecasting by learning the small-scale nature. arXiv. <https://arxiv.org/abs/2505.19432>
 110. Valcke S., Balaji V., Craig A., et al. (2012). Coupling technologies for Earth System Modelling. *Geosci. Model Dev.* **5**:1589–1596. DOI:10.5194/gmd-5-1589-2012
 111. Wilkinson M. D., Dumontier M., Aalbersberg I. J., et al. (2016). The FAIR guiding principles for scientific data management and stewardship. *Sci. Data* **3**:160018. DOI:10.1038/sdata.2016.18
 112. Liu Z. and Yao T. (2024). Strategizing Earth science data development. *Sci. Data* **11**:693. DOI:10.1038/s41597-024-03531-6
 113. Rödder S., Heymann M. and Stevens B. (2020). Historical, philosophical, and sociological perspectives on Earth system modeling. *J. Adv. Model. Earth Syst.* **12**:e2020MS002139. DOI:10.1029/2020MS002139
 114. Beutler T. (2021). Enforcing analytic constraints in neural networks emulating physical systems. *Phys. Rev. Lett.* **126**:098302. DOI:10.1103/PhysRevLett.126.098302
 115. Pegion K., Kirtman B. P., Becker E., et al. (2019). The subseasonal experiment (SubX): A multimodel subseasonal prediction experiment. *B. Am. Meteorol. Soc.* **100**:2043–2060. DOI:10.1175/BAMS-D-18-0270.1
 116. Raissi M., Perdikaris P. and Karniadakis G. E. (2019). Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.* **378**:686–707. DOI:10.1016/j.jcp.2018.10.045
 117. Hansen D., Maddix D.C., Alizadeh S., et al. (2024). Learning physical models that can respect conservation laws. *Physica D* **457**:133952. DOI:10.1016/j.physd.2023.133952
 118. Lee J., Gleckler P.J., Ahn M.S., et al. (2024). Systematic and objective evaluation of Earth system models. *Geosci. Model Dev.* **17**:3919–3948. DOI:10.5194/gmd-17-3919-2024
 119. Eyring V., Gleckler P. J., Heinze C., et al. (2016). Towards improved and more routine Earth system model evaluation in CMIP. *Earth Syst. Dyn.* **7**:813–830. DOI:10.5194/esd-7-813-2016
 120. Runge J., Gerhardus A., Varando G., et al. (2023). Causal inference for time series. *Nat. Rev. Earth Environ.* **4**:487–505. DOI:10.1038/s43017-023-00431-y
 121. Runge J., Bathiany S., Bolt E., et al. (2019). Inferring causation from time series in Earth system sciences. *Nat. Commun.* **10**:2553. DOI:10.1038/s41467-019-10105-3
 122. Govett M., Bah B., Bauer, P., et al. (2024). Exascale computing and data handling: Challenges and opportunities for weather and climate prediction. *Bull. Am. Meteorol. Soc.* **105**:E2385–E2404. DOI:10.1175/BAMS-D-23-0220.1
 123. Luccioni A. S., Viguer S. and Ligozat A.L. (2023). Estimating the carbon footprint of BLOOM, a 176B parameter language model. *J. Mach. Learn. Res.* **24**:1–15.
 124. Rozet F. and Louppe G. (2023). Score-based data assimilation. Oh A., Naumann T., Globerson A., et al. (eds). Advances in Neural Information Processing Systems 36 (Curran Associates), pp: 40521–40541. DOI: 10.52202/075280-1763
 125. Tom G., Schmid S. P., Baird S. G., et al. (2024). Self-driving laboratories for chemistry and materials science. *Chem. Rev.* **124**:9633–9732. DOI:10.1021/acs.chemrev.4c00055
 126. Rußwurm M., Wang S., Kellenberger B., et al. (2024). Meta-learning to address diverse earth observation problems across resolutions. *Commun. Earth Environ.* **5**:37. DOI:10.1038/s43247-023-01146-0
 127. Batzner S., Musaelian A., Sun L., et al. (2022). E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nat. Commun.* **13**:2453. DOI:10.1038/s41467-022-29939-5
 128. Shan S., Wang P., Chen S., et al. (2024). PiRD: Physics-informed residual diffusion for flow field reconstruction. arXiv. <https://arxiv.org/abs/2404.08412v2>
 129. Raskin R. G. and Pan M. J. (2005). Knowledge representation in the semantic web for Earth and environmental terminology (SWEET). *Comput. Geosci.* **31**:1119–1125. DOI:10.1016/j.cageo.2004.12.004
 130. Richard S.M. and CGI Interoperability Working Group (2007). GeoSciML – A GML application for geoscience information interchange. Solter D.R. (ed). Digital Mapping Techniques'06–Workshop Proceedings (U.S. Geological Survey), pp: 47–59. <https://pubs.usgs.gov/of/2007/1285/pdf/Richard.pdf>
 131. Wang C., Ma X., Chen J., et al. (2018). Information extraction and knowledge graph construction from geoscience literature. *Comput. Geosci.* **112**:112–120. DOI:10.1016/j.cageo.2017.12.007
 132. Zhang X., Huang Y., Zhang C., et al. (2022). Geoscience knowledge graph (GeoKG): development, construction and challenges. *Trans. GIS* **26**:2480–2494. DOI:10.1111/tgis.12985
 133. Zhou C., Wang H., Wang C., et al. (2021). Geoscience knowledge graph in the big data era. *Sci. China Earth Sci.* **64**:1105–1114. DOI:10.1007/s11430-020-9750-4
 134. Zhu Y., Sun K., Wang S., et al. (2023). An adaptive representation model for geoscience knowledge graphs considering complex spatiotemporal features and relationships. *Sci. China Earth Sci.* **66**:2563–2578. DOI:10.1007/s11430-022-1169-9
 135. Ho J., Jain A. and Abbeel P. (2020). Denoising diffusion probabilistic models. arXiv. <https://arxiv.org/abs/2006.11239v2>
 136. Schölkopf B., Locatello F., Bauer S., et al. (2021). Towards causal representation learning. arXiv. <https://arxiv.org/abs/2102.11107>
 137. Runge J., Nowack P., Kretschmer M., et al. (2019). Detecting and quantifying causal associations in large nonlinear time series datasets. *Sci. Adv.* **5**:eaau4996. DOI:10.1126/sciadv.aau4996
 138. Bareinboim E. and Pearl J. (2016). Causal inference and the data-fusion problem. *Proc. Natl. Acad. Sci.* **113**:7345–7352. DOI:10.1073/pnas.1510507113
 139. Wei J., Yang Y., Zhang X., et al. (2025). From AI for science to agentic science: a survey on autonomous scientific discovery. arXiv. <https://arxiv.org/abs/2508.14111>
 140. Bell A., Aides A., Helmy A., et al. (2026). Earth AI: Unlocking geospatial insights with foundation models and cross-modal reasoning. arXiv. <https://arxiv.org/abs/2510.18318>
 141. Feng P., Lv Z., Ye J., et al. (2026). Earth-agent: unlocking the full landscape of Earth observation with agents. *Int. Conf. Learn. Represent.* **2026**:55701–55764.
 142. Zhao S., Liu F., Zhang X., et al. (2026). OpenEarth-agent: from tool calling to tool creation for open-environment earth observation. arXiv. <https://arxiv.org/abs/2603.22148>
 143. Coley C. W., Thomas D. A., Lummiss J. A. M., et al. (2019). A robotic platform for flow synthesis of organic compounds informed by AI planning. *Science* **365**:eaax1566. DOI:10.1126/science.aax1566
 144. Shields B. J., Stevens J., Li J., et al. (2021). Bayesian reaction optimization as a tool for chemical synthesis. *Nature* **590**:89–96. DOI:10.1038/s41586-021-03213-y
 145. King R. D., Rowland J., Oliver S. G., et al. (2009). The automation of science. *Science* **324**:85–89. DOI:10.1126/science.1165620

FUNDING AND ACKNOWLEDGMENTS

This work was funded by the Strategic Priority Research Program of Chinese Academy of Sciences (XDB0740300), the National Natural Science Foundation of China (42501457), and the Innovation Project of LREIS (YKPI507). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

AUTHOR CONTRIBUTIONS

F. S. and Y. C. conceived the review and designed its overall framework. Y. C. led the literature review, synthesis of the relevant studies, preparation of the figures, and drafting of the manuscript. F. S. supervised the work and contributed to the conceptual development, scientific interpretation, and critical revision of the manuscript. X. L., H. W., R. F., S. H., W. H., Z. C., P. Z., L. Z., T. A., G. H., H. T., X. L., Z. Z., C. D., Z. L., L. W., J. W., and C. J. contributed to the literature review, scientific discussion, interpretation, and revision of the manuscript. All authors contributed to the manuscript and approved the final version.

DECLARATION OF INTERESTS

The authors declare no competing interests.