From self-supervision to trustworthy EO foundation models

DARES 2025

Ioannis Papoutsis (ipapoutsis@mail.ntua.gr)

Assistant Professor, National Technical University of Athens

Associate Researcher, National Observatory of Athens

Affiliated Researcher, Archimedes / Athena Research Center

https://orion-ai-lab.github.io/

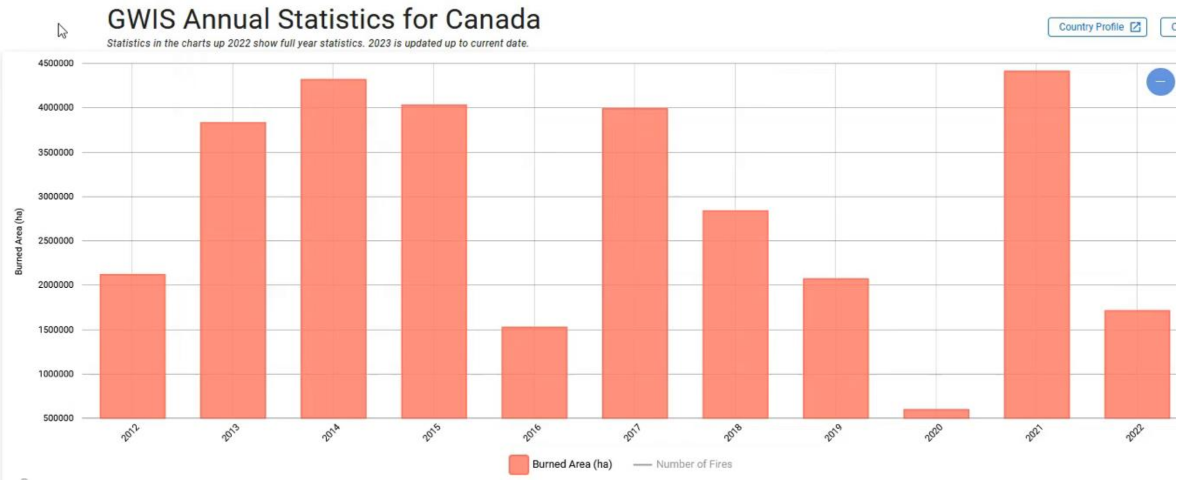
RSLab
Remote Sensing Laboratory
National Technical University of Athens

orion lab
AI & Earth Observation Research

Forecasting wildfires

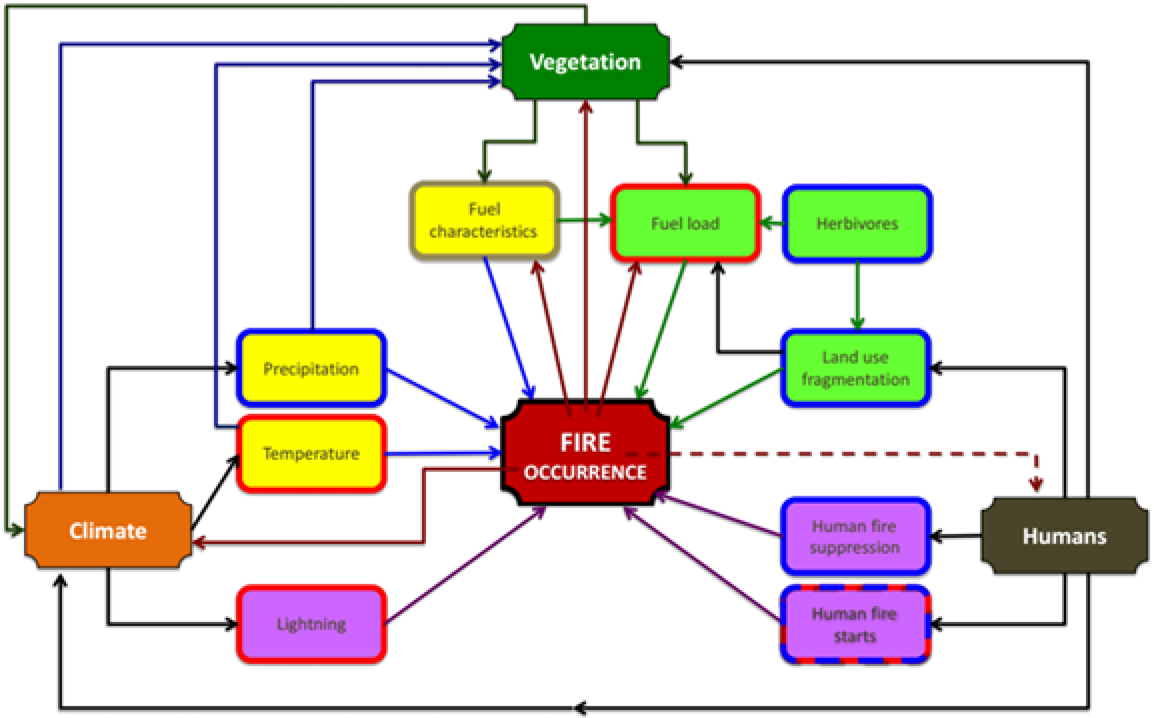
Motivation

- High variability between fire seasons
- Climate change fosters extreme fire conditions



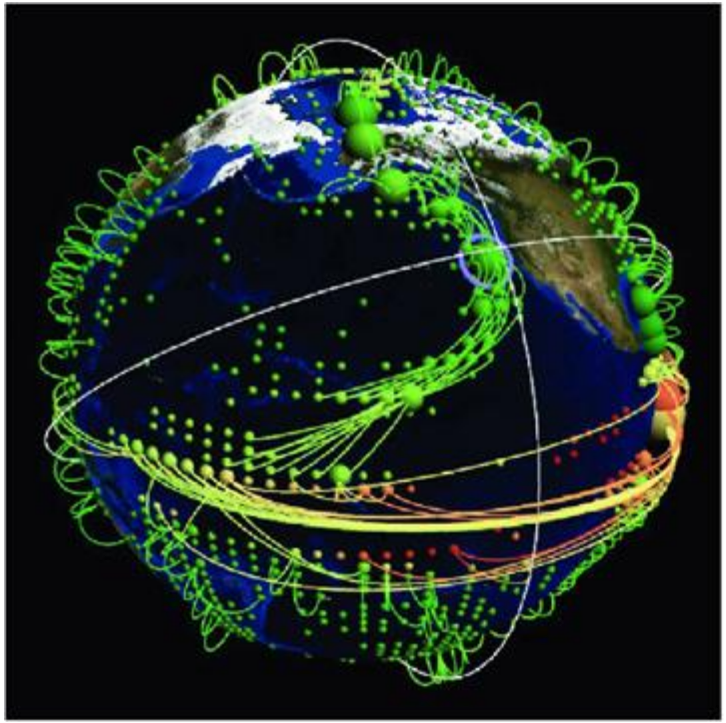
Prapas et al., Deep Learning Methods for Daily Wildfire Danger Forecasting. *NeurIPS Workshop on Artificial Intelligence for Humanitarian Assistance and Disaster Response* (2021)

Challenges



Fire Drivers. *Source: Hantson et al. "The status and challenge of global fire modelling" (2016)*

Earth is a complex inter-connected system



Source: Statistical physics approaches to the complex Earth system

- **Teleconnections** are long-range spatiotemporal connections in the earth system
- **Memory effects** refer to the temporal dynamics of earth system variables

nature communications

Article https://doi.org/10.1038/s41467-023-36052-8

Climate teleconnections modulate global burned area

Received: 31 March 2022

Accepted: 12 January 2023

Adrián Cardil[1,2,3], Marcos Rodrigues[4,5], Mario Tapia[2], Renaud Barbero[6], Joaquin Ramírez[2], Cathelijne R. Stoof[7], Carlos Alberto Silva[8], Midhun Mohan[9] & Sergio de-Miguel[1,3]

Chen et al. (2016), Env. Res. Letts.

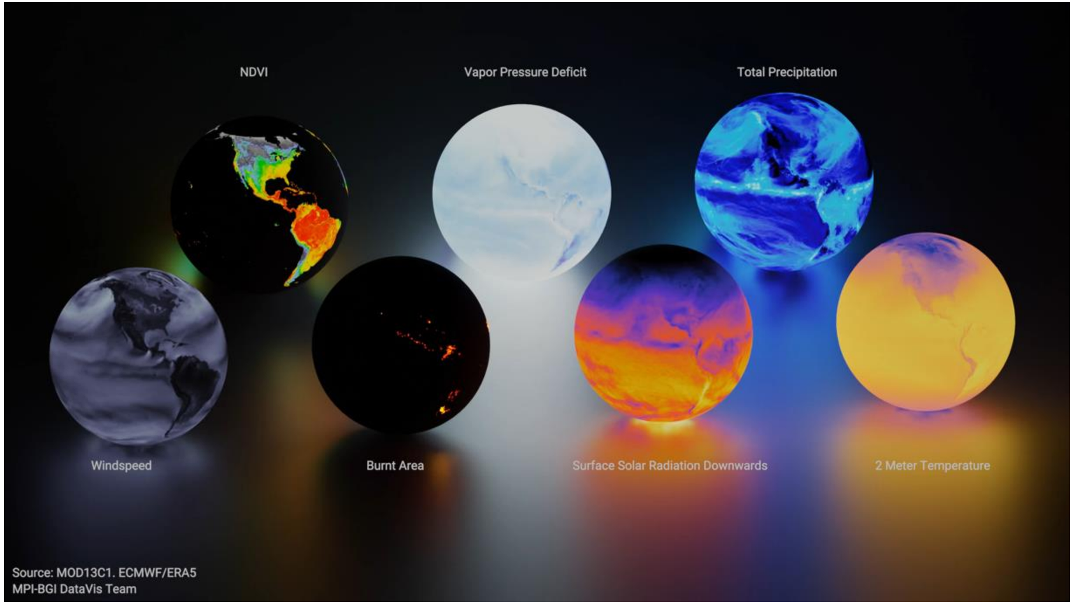
Kim et al. (2020), Sci. Adv.

Yu et al. (2020), Nature coms.

Justino et al. (2022), Clim. & Atm. Sci.

Cardil et al. (2023), Nature coms.

SeasFire Cube



SeasFire Cube: A Global Dataset for Seasonal Fire Modeling in the Earth System [Data set]. *Zenodo*. https://doi.org/10.5281/zenodo.7108392

Karasante et al. SeasFire cube-a multivariate dataset for global wildfire modeling. *Scientific Data* (2025)

Design choices: granularity

Resolution: 8days x 0.25° x 0.25°

Extent: Global, 2001 - 2020

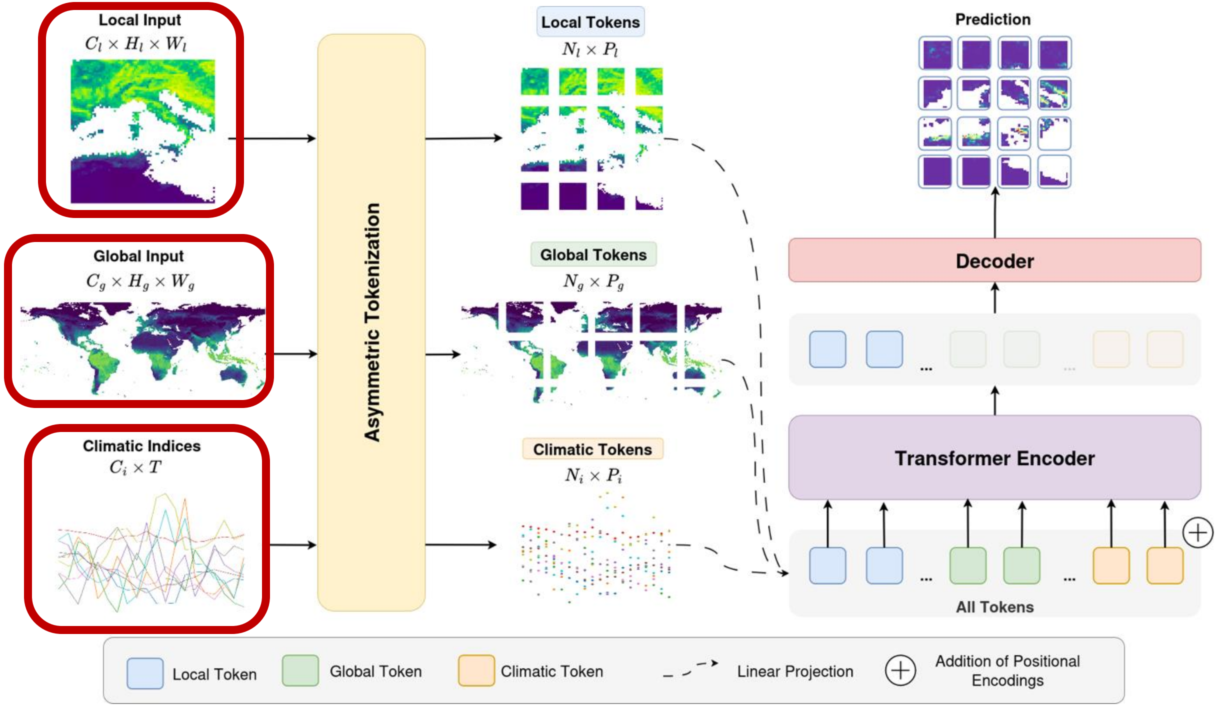
Wildfire drivers

- Meteorology (ERA5)
- Satellite Observations (MODIS)
- Vegetation, Surface Temperature
- Oceanic Indices (NOAA)
- Population Density (NASA SEDAC), Land Cover (ESA CCI)

Wildfire variables

- Burned Areas (GFED, FireCCI, GWIS)
- Fire Emissions (GFAS)

Transformers: TeleViT architecture

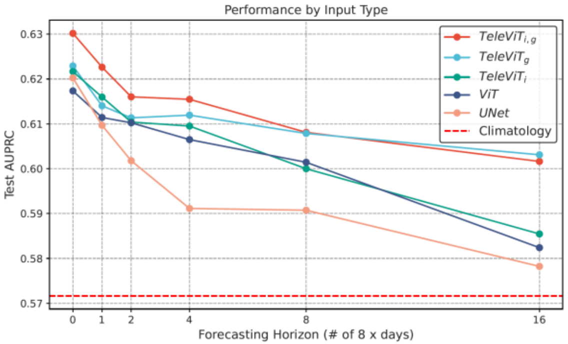


TeleViT: Teleconnection-driven Transformers Improve Subseasonal to Seasonal Wildfire Forecasting

Ioannis Prapas (1, 3), Nikolaos-Ioannis Bountos (1, 2), Spyros Kondylatos (1, 3), Dimitrios Michail (2), Gustau Camps-Valls(3), Ioannis Papoutsis (1)
(1) Orion Lab, IAASARS, National Observatory of Athens
(2) Department of Informatics and Telematics, Harokopio University of Athens
(3) Image & Signal Processing Group, Universitat de València

Best Paper Award at ICCV 2023, AI+HADR workshop

Paper | Code | arXiv



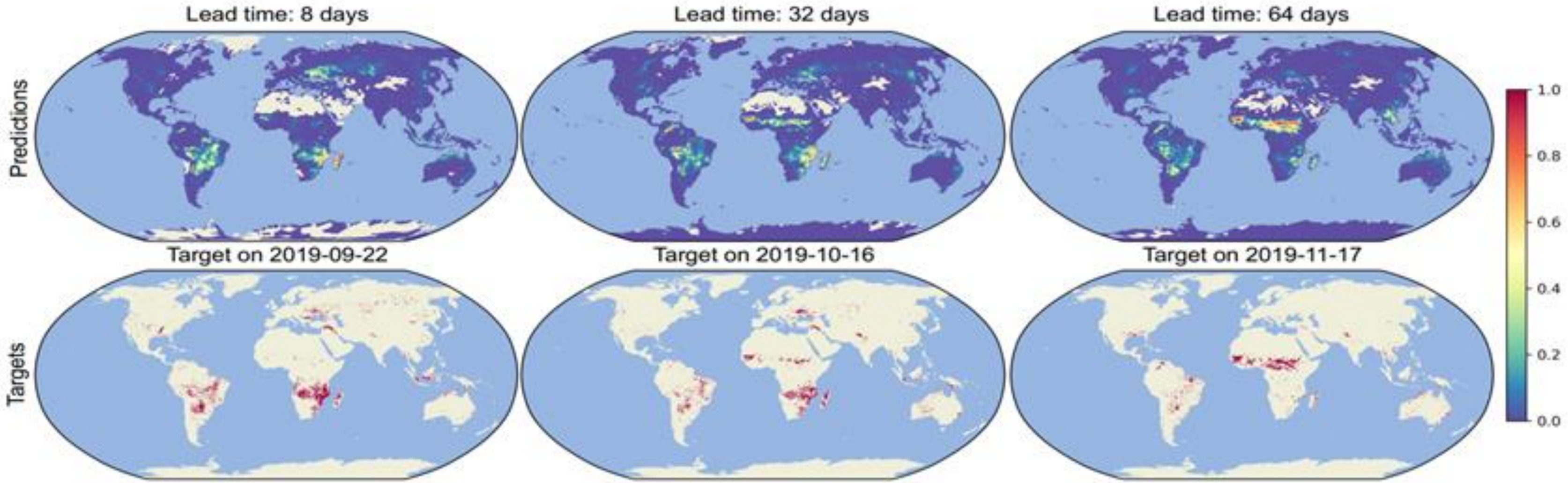
How robust are the predictions?

TeleViT: Teleconnection-driven Transformers Improve Subseasonal to Seasonal Wildfire Forecasting

Ioannis Prapas (1, 3), Nikolaos-Ioannis Bountos (1, 2), Spyros Kondylatos (1, 3), Dimitrios Michail (2), Gustau Camps-Valls(3), Ioannis Papoutsis (1)
(1) Orion Lab, IAASARS, National Observatory of Athens
(2) Department of Informatics and Telematics, Harokopio University of Athens
(3) Image & Signal Processing Group, Universitat de València
Best Paper Award at ICCV 2023, AI+HADR workshop

Paper Code arXiv

ICCV23 PARIS



ecai 2025 BOLOGNA

RSLab Remote Sensing Laboratory National Technical University of Athens

orion lab

Michail et al. FireCastNet: Earth-as-a-graph for seasonal fire prediction. arXiv preprint arXiv (2025)

How robust are the predictions?

GFED Region	Fraction of burned areas	Horizon					
		1	2	4	8	16	24
Boreal North America (BONA)	0.924%	0.04	0.03	0.03	0.03	0.02	0.02
Temperate North America (TENA)	1.986%	0.26	0.27	0.26	0.26	0.25	0.26
Central America (CEAM)	2.137%	0.55	0.54	0.53	0.53	0.55	0.55
Northern Hemisphere South America (NHSA)	2.673%	0.67	0.65	0.64	0.63	0.61	0.61
Southern Hemisphere South America (SHSA)	15.619%	0.44	0.43	0.43	0.42	0.41	0.41
Europe (EURO)	0.857%	0.20	0.20	0.19	0.18	0.17	0.19
Middle East (MIDE)	1.029%	0.29	0.30	0.29	0.30	0.29	0.26
Northern Hemisphere Africa (NHAF)	20.749%	0.74	0.73	0.72	0.71	0.71	0.72
Southern Hemisphere Africa (SHAF)	29.988%	0.85	0.84	0.84	0.83	0.83	0.84
Boreal Asia (BOAS)	4.072%	0.14	0.13	0.13	0.13	0.13	0.14
Central Asia (CEAS)	8.264%	0.27	0.26	0.27	0.26	0.25	0.24
Southeast Asia (SEAS)	5.764%	0.64	0.63	0.62	0.61	0.60	0.61
Equatorial Asia (EQAS)	1.089%	0.49	0.50	0.48	0.44	0.40	0.40
Australia and New Zealand (AUST)	4.849%	0.31	0.31	0.30	0.30	0.29	0.32

Class imbalance

A blessing and a curse - Class imbalance in natural hazards

AI4Extremes

- Natural hazards are by definition extreme events → **Rare**
- Difficult to acquire a dedicated dataset for each problem
  - Depending on the problem formulation, the spatial coverage and the used **sensors**, different annotations may be required
  - Annotations require expert knowledge
- **Spatiotemporal generalization** becomes way harder with limited data
  - E.g. It is extremely difficult to predict burned areas in Africa, when using data solely from the Mediterranean for training

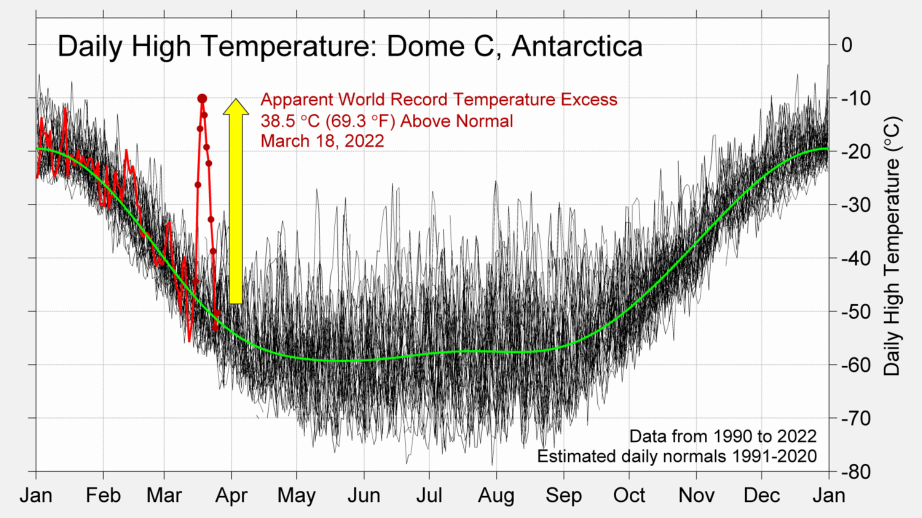


Image source: Dr. Robert Rohde

AI4EO challenges for disaster management

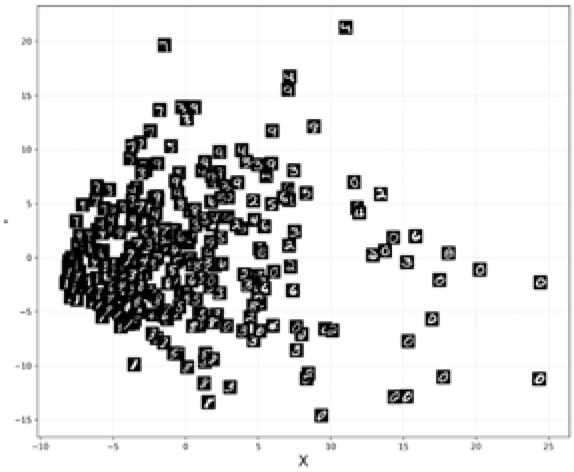
- Big data
- Labeling (imbalance + noisy labels)
- Generalization
- Uncertainty in forecasting
- Stochastic nature, complex, non-linear

Large unlabeled datasets

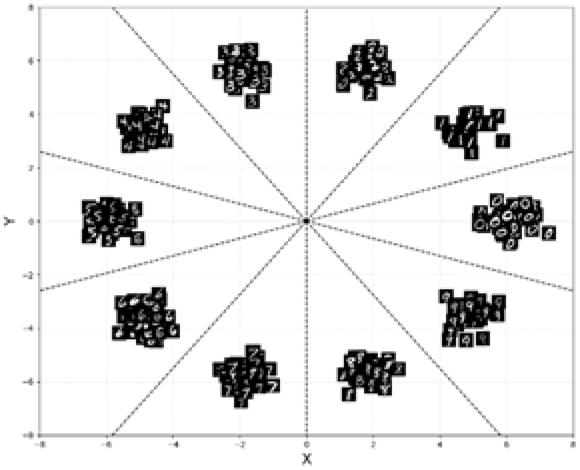
Self-supervised pre-training

Representation learning

- Deep learning success depends on learning meaningful, information-rich representations by training on large datasets
- Images are complex high-dimensional arrays → What is a good representation?



(a) 2d projection of MNIST using PCA



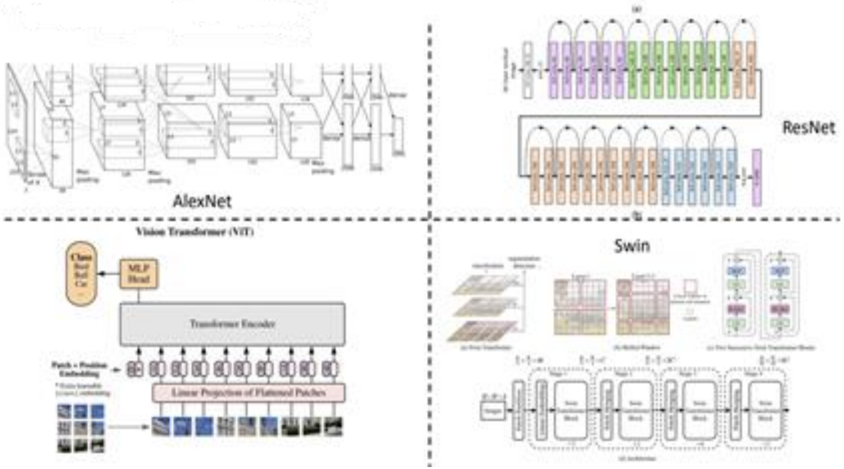
(b) Ideal 2d representation space

Supervised learning

Mapping an image to a discrete label which is associated to a visual concept

- Early/powerful foundation models when trained on large datasets
- Standard way to develop model backbones

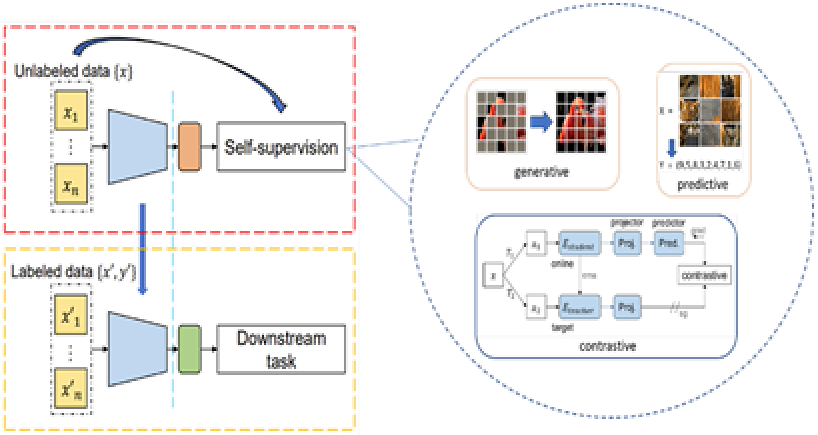
- Annotation is expensive and limited!
  - Especially in EO where expert eyes or field measurements might be needed



Gan, Zhe, et al. "Recent advances in vision foundation models" *CVPR 2023 tutorial.*

SSL in the natural image domain

- Solving a meaningful pretext task
  - e.g. Rotation prediction and Jigsaw puzzles
- Instance discrimination
  - Contrastive learning e.g., SimCLR, MoCo, etc.
- Information restoration
  - e.g. Colorization, Masked Autoencoders, etc.
- Self-distillation methods
  - e.g. BYOL, SimSiam



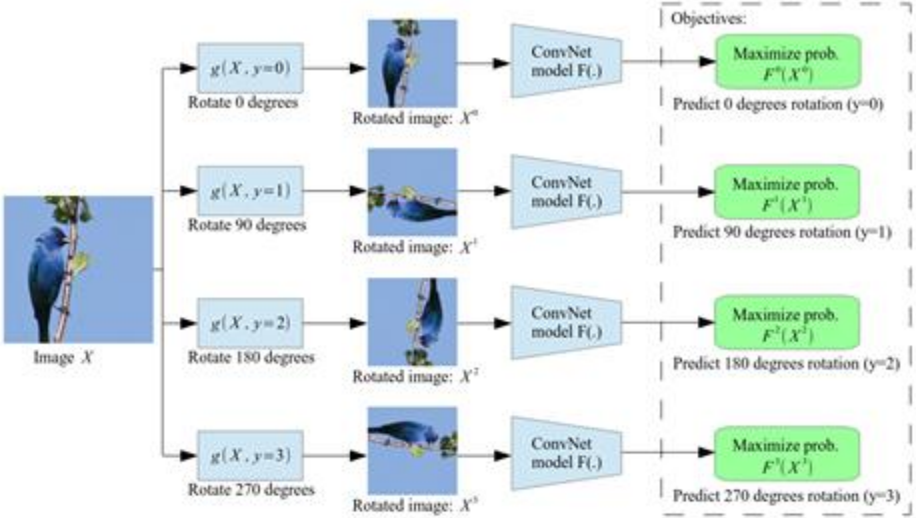
Wang, Yi, et al. "Self-Supervised Learning in Remote Sensing: A Review." *IEEE GRSM* (2022).

Nice overview on SSL methods and tips:
Balestriero et al. A cookbook of self-supervised learning. *arXiv* (2023).

A. Predictive self-supervised learning

Hand-designed pretext tasks utilizing the intrinsic characteristics of data

- Solving Jigsaw puzzles [1]
- Rotation prediction [2]



[1] Noroozi et al., Unsupervised learning of visual representations by solving jigsaw puzzles. *European conference on computer vision* (2016)
[2] Gidaris et al., Unsupervised representation learning by predicting image rotations. *arXiv* (2018)

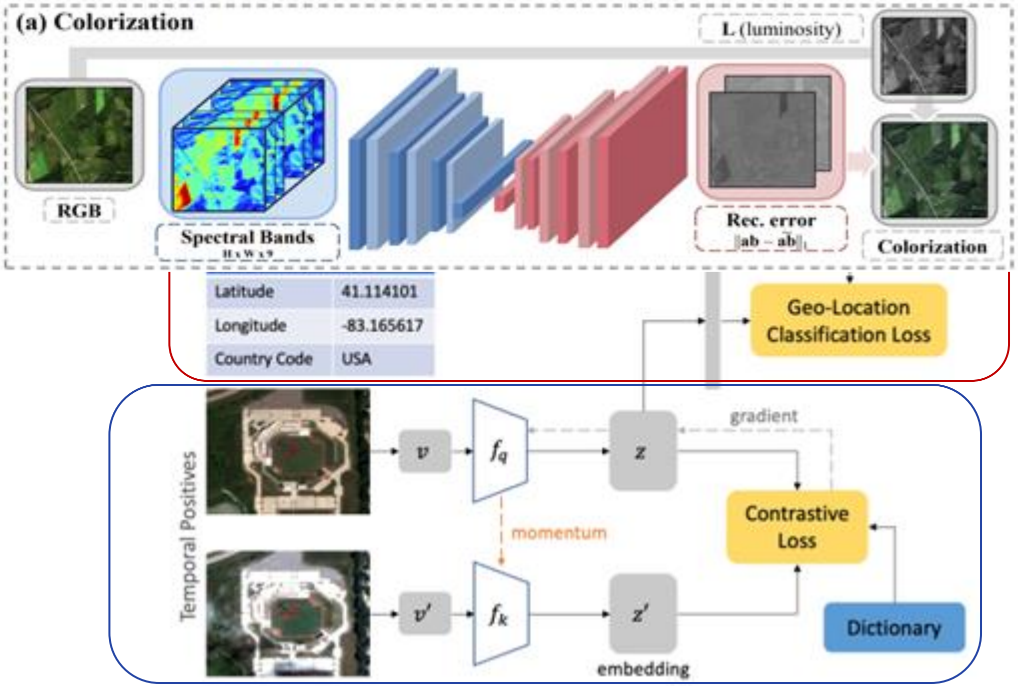
A. Predictive self-supervised learning in EO

The color out of space[1]:

- Predict RGB channels from multi-spectral information as a pretext task
- Downside: Learns an encoder only for MS channels, not RGB

Geography-aware SSL [2]

- Predict geolocation (grouped in K clusters) of an image
- MoCo-v2 formulation except the **positive pairs come from different temporal points**
- Intuitively, could provide invariance in temporal-based changes

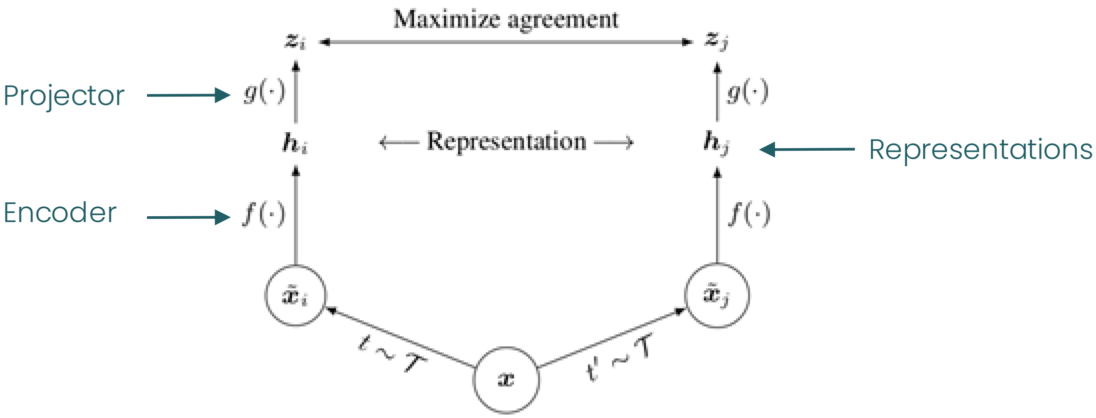


[1] Vincenzi et al., The color out of space: learning self-supervised representations for earth observation imagery. *ICPR IEEE* (2021)
[2] Ayush et al., Geography-aware self-supervised learning. *ICCV* (2021)

B. Contrastive self-supervised learning

Joint-embedding architectures with instance discrimination

- SimCLR [1]



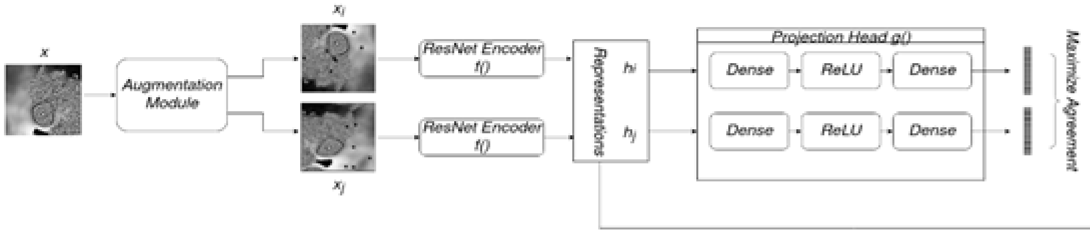
- Heavy dependance on augmentation set
- Heuristics from natural images don’t fit to other domains
- Negative samples are crucial to learning good representations
- Requires large batch sizes → More negative samples

[1] Chen et al., A simple framework for contrastive learning of visual representations. *ICML* (2020)

B. Contrastive self-supervised learning in EO

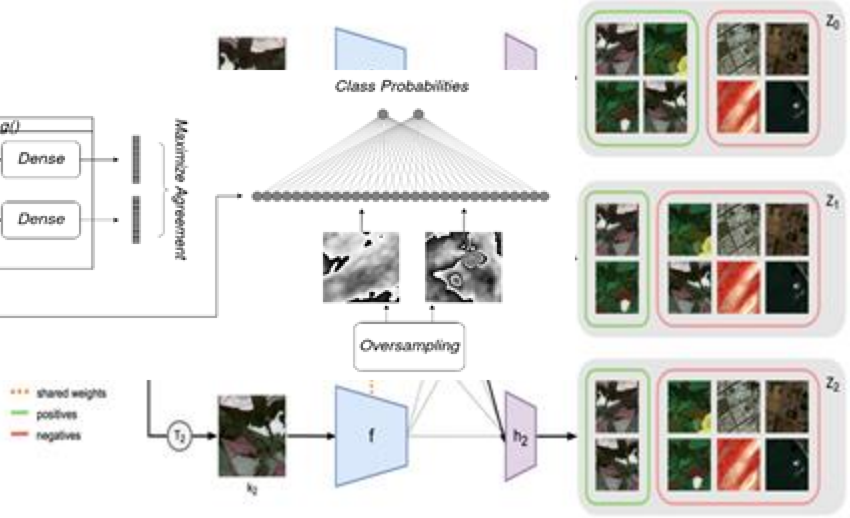
What makes for good views / data augmentations in EO?

- Tailored augmentations per sensor/product, e.g. for InSAR [1]



- Difficult to identify optimal set of augmentations per modality
- EO is highly multi-modal

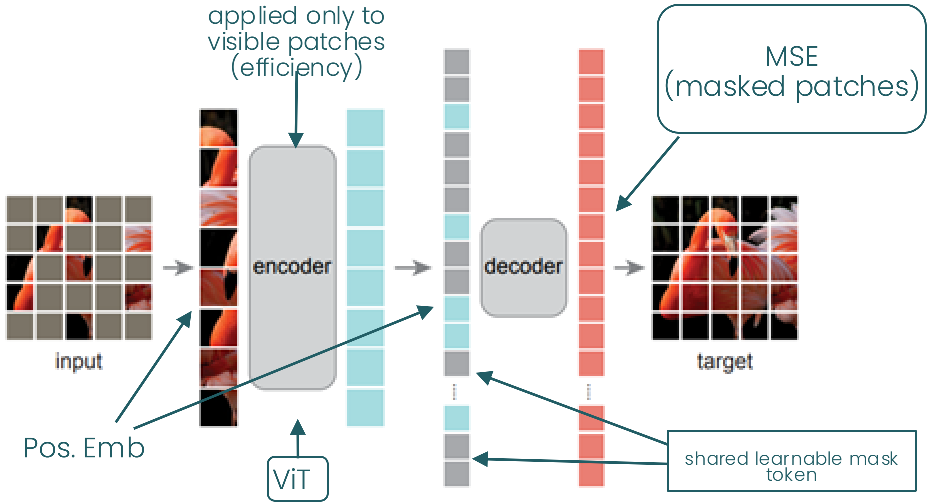
Three representation spaces



[1] Bountos, et al. Self-supervised contrastive learning for volcanic unrest detection. *IEEE Geoscience and Remote Sensing Letters* (2021)
[2] Manas, et al. Seasonal contrast: Unsupervised pre-training from uncurated remote sensing data. *ICCV* (2021)

C. Generative self-supervised learning

Masked Autoencoders [1] → Information restoration pretext task



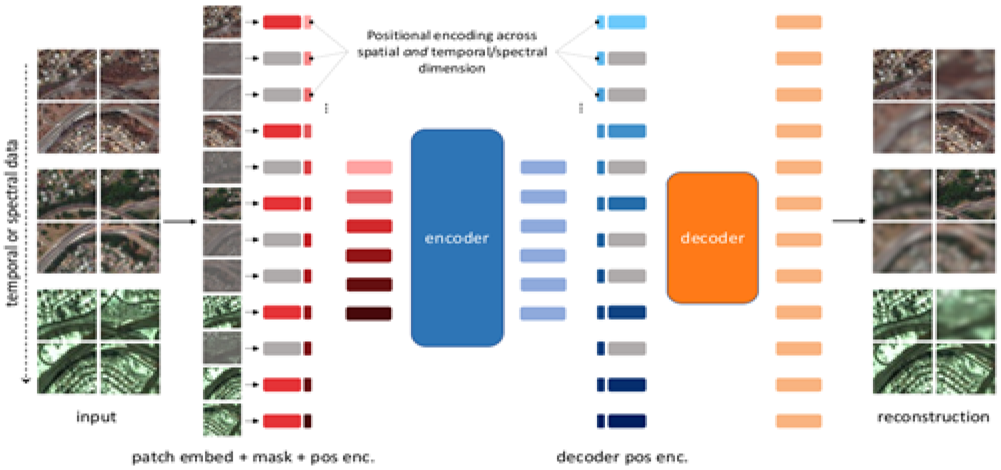
- Efficient
- Generic pretext task, applicable to any domain → highly popular in EO
- Does not rely on hand crafted augmentations → important to EO
- Good as initialization
- Representations are not highly discriminative[1][2] → Linear probing is ineffective

[1] He et al., Masked autoencoders are scalable vision learners. *CVPR* (2022)
[2] Przewięźlikowski et al., Beyond [cls]: Exploring the true potential of Masked Image Modeling representations. *arXiv* (2024)

C. Generative self-supervised learning for EO

SatMAE - Masked Autoencoder for multi-spectral temporal data [1]

- Carefully exploits the unique information from the temporal and spectral dimensions of RS data
- Patches can be generated in both temporal and spectral dimensions
- Temporal or spectral group encoding



[1] Cong et al., Satmae: Pre-training transformers for temporal and multi-spectral satellite imagery. *Advances in NeurIPS* (2022)

Overall guidelines

- Unless one has massive compute resources, MoCo-v2 is a good way to compensate for smaller batch sizes in contrastive learning
- Data augmentations should be carefully selected taking in mind the problem at hand
  - Considering temporal based augmentation is a convenient idea in Remote Sensing (free, naturally occurring data augmentations)
- MAE is a good way to avoid choosing an augmentation set
  - An ablation on the masking-ratio however, is necessary depending on the data source and the expected downstream tasks
- In real life scenarios we are not restricted to linear evaluation
  - Fine-tuning more layers may provide significant improvements

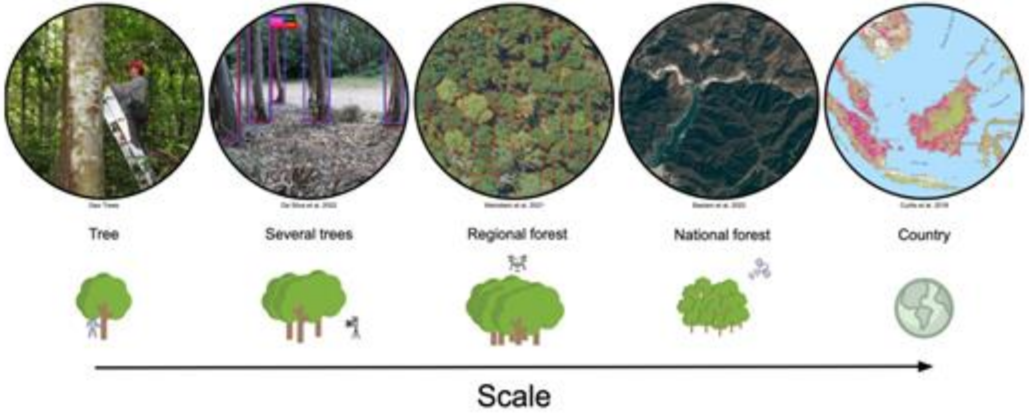
Foundation Models in EO

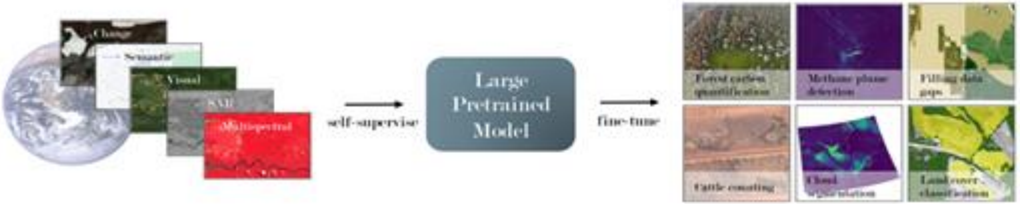
Sensor-agnostic FMs for EO

Earth Observation data vary in terms of:

- Scales of objects [1]
- Sensor
- GSD
  - From <1cm per pixel to >1km per pixel
- Environmental conditions

Vision: Can we create a generalist EO Foundation Model [2]?





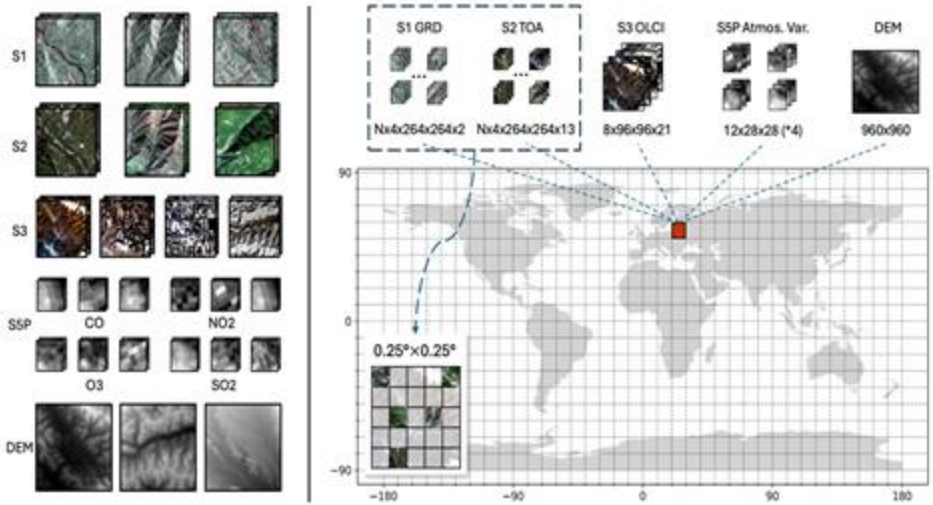
[1] Ouaknine et al., OpenForest: a data catalog for machine learning in forest monitoring. *Environmental Data Science* (2025)
[2] Lacoste et al., Geo-bench: Toward foundation models for earth monitoring. *Advances in NeurIPS* (2023):

EO pretraining datasets

Datasets getting bigger and bigger, when will it stop?

Dataset	Modality	Resolution	# Time stamps	# patches	# pixels
fMoW [14]	RGB, MS	0.3–10 m	3	2M	50B
SEN12MS [53]	SAR, MS	10 m	1	540K	35B
SeCo [42]	MS	10 m	5	1M	70B
SSL4EO-S12 [62]	SAR, MS	10 m	4	3M	140B
SSL4EO-L [54]	MS	30 m	4	5M	348B
SatlasPretrain [6]	SAR, MS, RGB	0.5–10 m	~10	>10M	17T
MMEarth [45]	SAR, MS, height, landcover, etc.	10–15 m	1	6M	120B
SpectralEarth [12]	HS	30 m	1–23	540K	10B
Major TOM [23]	SAR, MS	10 m	1	8M	6.8T
Copernicus-Pretrain	SAR, MS, S3, DEM, S5P	10 m–1 km	1–12	19M	920B

Wang et al., Towards a unified Copernicus foundation model for earth vision, *ICCV* (2025)



The evolution tree for EO foundation models

Model	Arch.	Pretrained EO Data	Learning Strategy	Params (M)	Year
CROMA	ViT	SSL4EO-S12 [25]	Contrastive	396.13	2023
DOFA	ViT	DOFA [21]	MIM	178.20	2024
GFM-Swin	Swin-T	GeoPile [22]	MIM	128.36	2023
Prithvi	ViT	Prithvi-HLS [23]	MIM	153.28	2023
RemoteClip	ViT	SEG-4, DET-10, RET-3 [19]	Contrastive	154.34	2024
SatlasNet	Swin-T	SatlasPretrain [26]	Supervised	128.57	2023
ScaleMAE	ViT	FMoW-RGB [48]	MIM	396.21	2023
SpectralGPT	ViT	fMoW-S2 [48], BigEarthNet [49]	MIM	614.75	2024
SSL4EO-S12	ViT	SSL4EO-S12 [25]	MIM	61.99	2022
SoftCon	ViT	SSL4EO-S12 [25]	Contrastive	242.19	2024

Wang et al., Towards a unified Copernicus foundation model for earth vision, *ICCV* (2025)

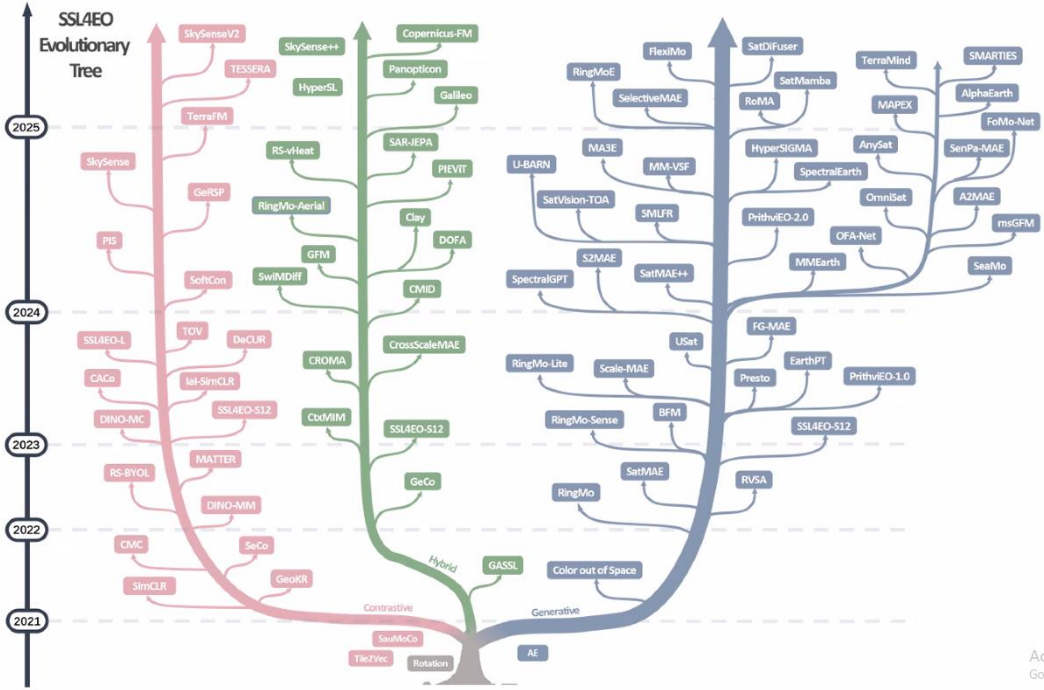


Image © Yi Wang. PhD Defense, Technical University of Munich (2025)

Requirements & evaluation of sensor-agnostic FMs

Zhu et al., On the Foundations of Earth and Climate Foundation Models. *arXiv* (2024)

To be effective, EO foundation models must:

- **Generalize** across diversity
  - Multi-sensor & multi-modality
  - Global, multi-scale, multi-temporal
- Be robust to **real-world** challenges
  - Resist spatio-temporal shifts
  - Handle data scarcity (limited coverage, costly labels, rare events)
- Produce **quality representations**
  - Meaningful per modality
  - Capture cross-modal interactions
- Be broadly applicable
  - **Flexible across tasks** (classification, segmentation, detection, change, instance)
  - **Benchmarked** on diverse, harmonized EO datasets

Name	# Tasks	RGB	Multispectral	Hyperspectral	SAR	Time Series	Classification	Regression	Segmentation	Object Det.	Change Det.	Resolution	Timespan
SustainBench [109]	15	✓	✓	✗	✗	✓	✓	✓	✓	✗	✗	0.6 m–30 m	1996–2019
GEO-Bench [110]	12	✓	✓	✓	✓	✗	✓	✗	✓	✗	✗	0.1 m–15 m	2001–2021
FoMo-Bench [111]	16	✓	✓	✓	✓	✗	✓	✗	✓	✓	✗	0.01 m–60 m	2011–2023
MDAS [112]	3	✓	✓	✓	✓	✗	✗	✗	✓	✗	✗	0.3 m–30 m	2018
PhilEO Bench [113]	3	✗	✓	✗	✗	✗	✗	✓	✓	✗	✗	10 m	unknown
SkySense [48]	17	✓	✓	✗	✓	✓	✓	✗	✓	✓	✓	0.05 m–30 m	2002–2019
Prithvi [114]	4	✗	✓	✗	✓	✓	✗	✗	✓	✗	✗	30 m	2017–2022
PANGAEA [115]	11	✓	✓	✗	✓	✓	✗	✓	✓	✗	✓	0.8 m–30 m	2015–2022

(a) Benchmarks for EO FMs.

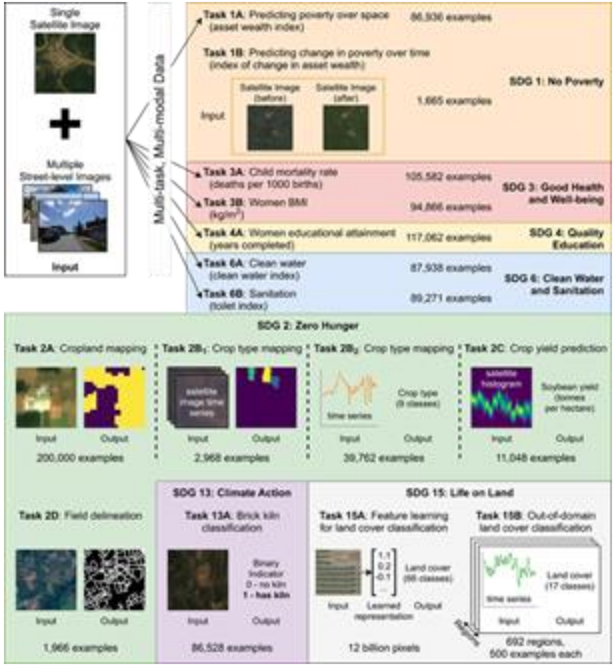
Name	Nowcasting	Medium-Range	S2S Forecasting	Climate Projection	Extreme Events	Downscaling	Resolution (Lat)	Resolution (Lon)	Resolution (Time)	Timespan
ERA5† [29]	✓	✓	✓	✓	✗	✗	0.25°	0.25°	1 h	1959–2024
WeatherBench [116]	✗	✓	✗	✗	✗	✗	5.63°	5.63°	6 h	1979–2018
WeatherBench 2 [117]	✗	✓	✗	✗	✗	✗	1.5°	1.5°	6 h	1979–2023
ExtremeWeather [118]	✗	✗	✗	✗	✓	✗	0.23°	0.31°	6 h	1979–2005
ClimateLearn [119]	✗	✓	✗	✗	✓	✓	5.63°	5.63°	6 h	1979–2018
CMIP6† [33]	✗	✓	✓	✓	✗	✗	1.25°	2.5°	24 h	1850–2100
ClimateBench [120]	✗	✗	✗	✓	✗	✗	1.89°	2.5°	1 yr	1850–2100
MRMS† [121]	✓	✓	✓	✗	✗	✓	0.01°	0.01°	2 min	2014–2024

†Raw dataset without associated evaluation protocol.

(b) Benchmarks for weather and climate FMs.

Evaluation of sensor-agnostic FMs

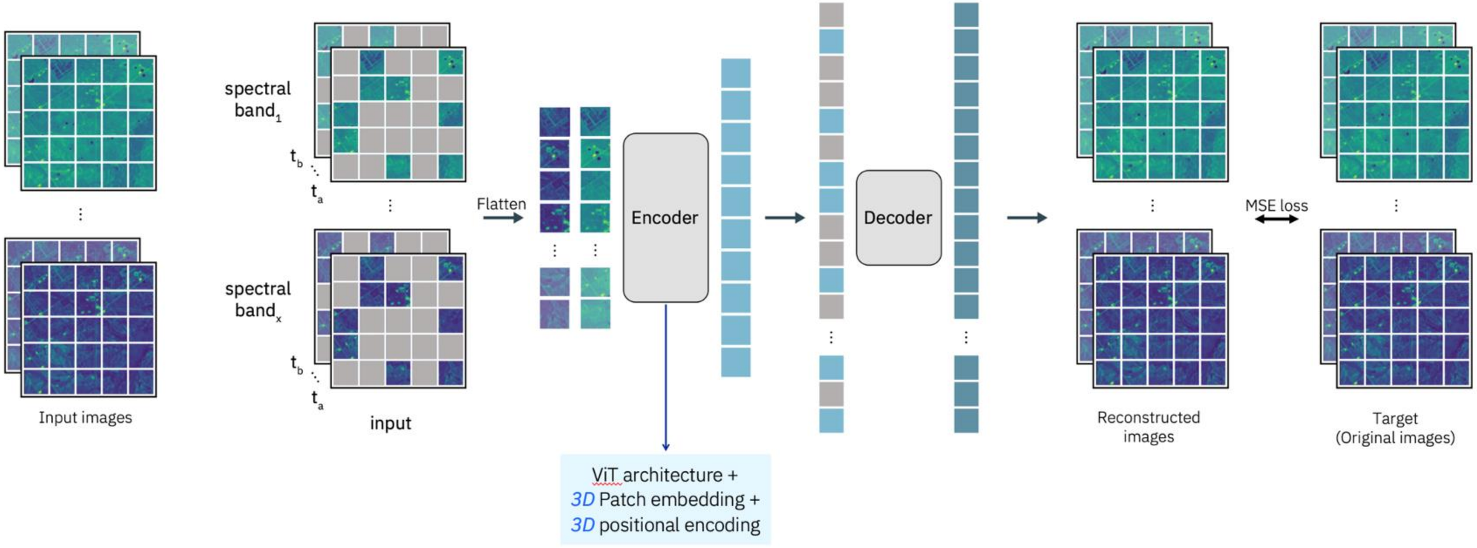
- Sustain-Bench [1]
  - 15 tasks oriented towards 7 SDGs
  - Multimodal
  - Global spatial coverage



[1] Yeh et al., Sustainbench: Benchmarks for monitoring the sustainable development goals with machine learning. *arXiv* (2021).

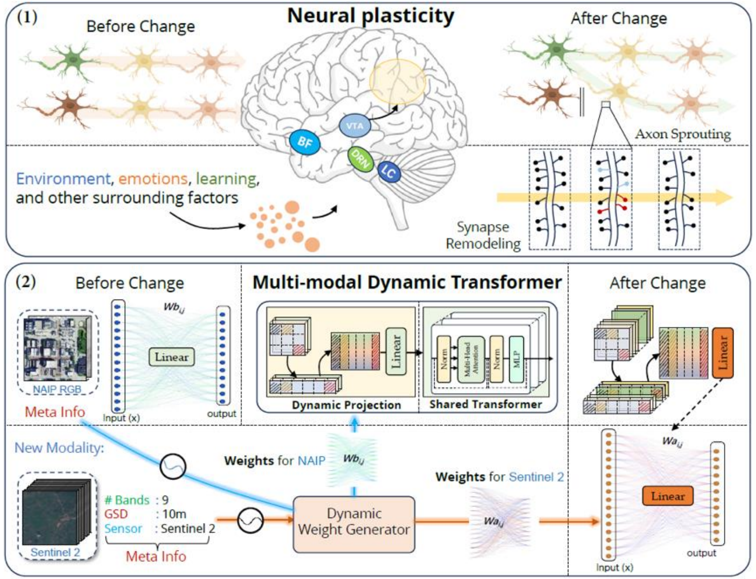
Unimodal FM – Prithvi 1.0

Based on the Harmonized Landsat-Sentinel (HLS) imagery

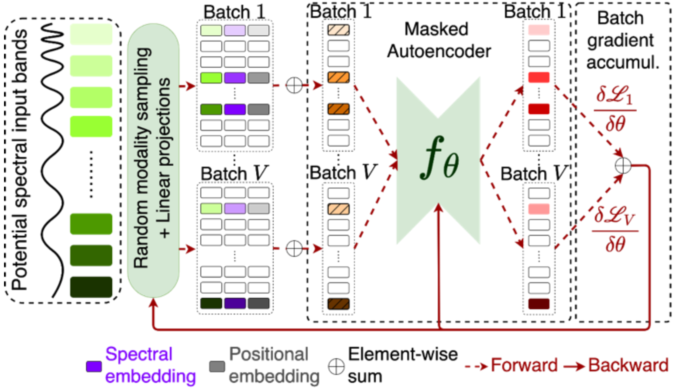


Jakubik et al., Foundation models for generalist geospatial artificial intelligence. *arXiv* (2023)

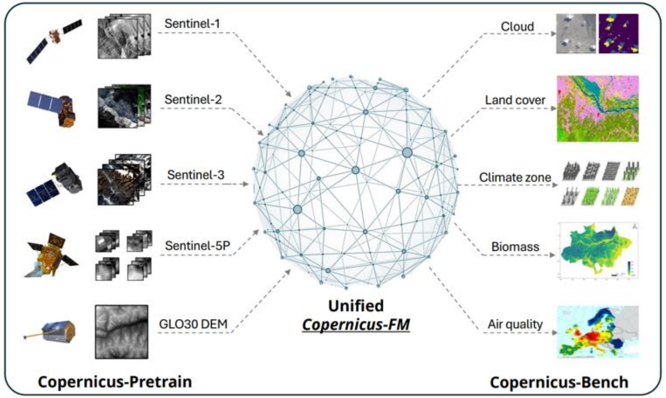
Challenges – multimodal



Xiong et al., Neural Plasticity-Inspired Multimodal Foundation Model for Earth Observation. *arXiv* (2024)



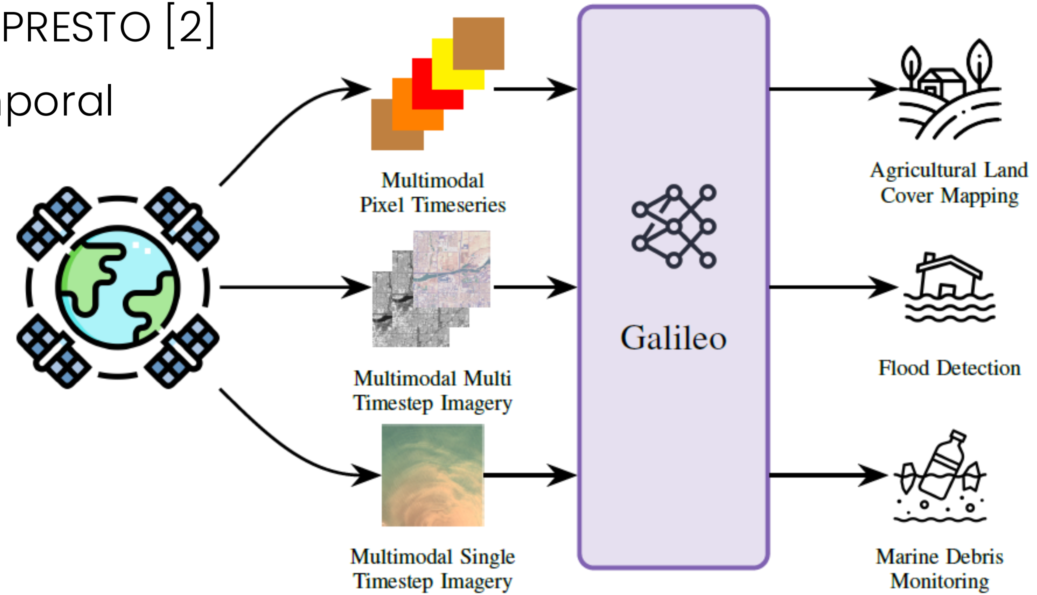
Bountos et al., FoMo: Multi-Modal, Multi-Scale and Multi-Task Remote Sensing Foundation Models for Forest Monitoring. *AAAI* (2025)



Yi et al., Towards a Unified Copernicus Foundation Model for Earth Vision. *ICCV* 2025

Challenges – temporal dimension

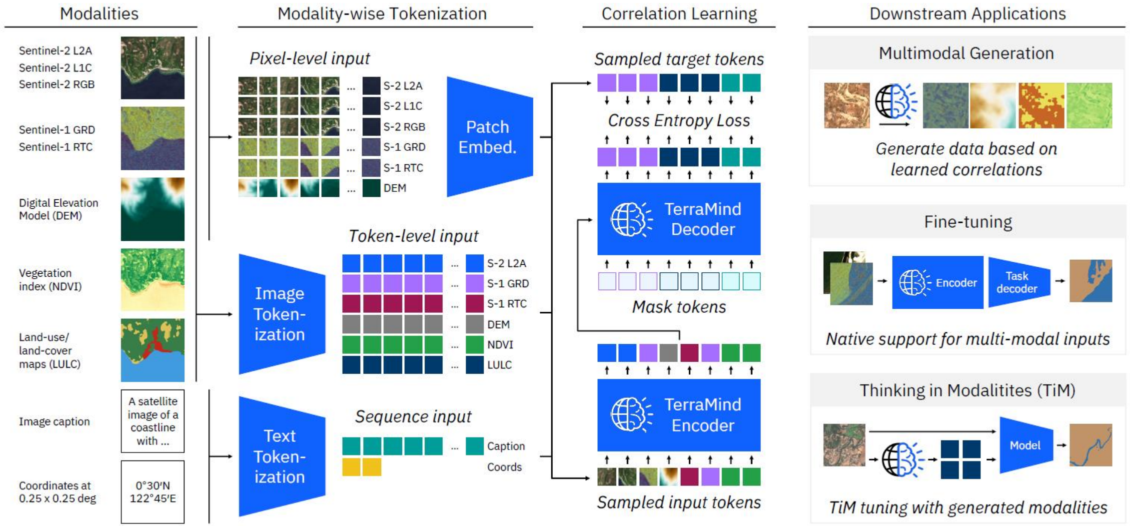
- Galileo [1], the successor of PRESTO [2]
- AnySat [3] is also multi-temporal



[1] Tseng et al., Galileo: Learning Global & Local Features of Many Remote Sensing Modalities. *arXiv* (2025)
[2] Tseng et al., Lightweight, pre-trained transformers for remote sensing timeseries. *arXiv* (2023)
[3] Astruc et al., AnySat: One Earth Observation Model for Many Resolutions, Scales, and Modalities. *CVPR* (2025)

Challenges - scale

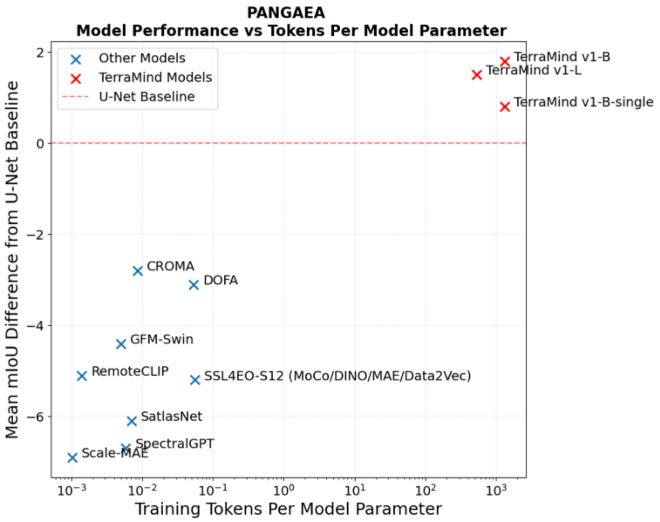
TerraMind is the first "**any-to-any**" generative, multimodal foundation model for EO, pre-trained on a massive dataset comprising **1 trillion tokens** derived from **9 million spatio-temporal samples** across **nine geospatial modalities**, including optical, radar, DEM, NDVI, land-use maps, coordinates, and captions

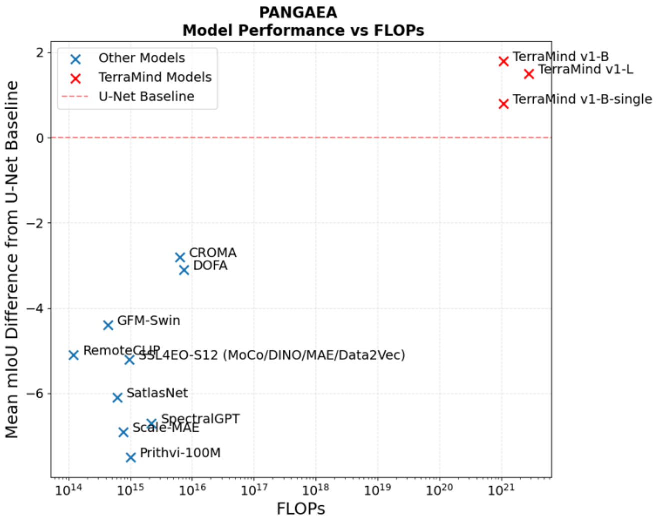


Jakubik et al., TerraMind: Large-Scale Generative Multimodality for Earth Observation. *arXiv* (2025)

Geospatial Foundational Disappointments

- https://christopherren.substack.com/p/geospatial-foundational-disappointments
- After 10^{21} FLOPs and 500 B patches, IBM's TerraMind beats a supervised U-Net by just +2 mIoU on PANGAEA; losing on 5/9 tasks, most other GFMs do worse
- Current pre-training objectives are unlikely to scale further with compute and data
- *"I am disappointed."*





Limitations

- There are many **shortcuts to achieve high performance in a dataset** but:
  - Can we assess **how well a model encodes intra- and inter-modality** properties and relationships?
- Pre-training is not harmonized:
  - **Difficult to identify the source of performance improvements**: Is it the pretraining dataset/setup or the methodology?
- Do we really need all these data? → Inherent **redundancy** in EO data
- Do we really need such **big models**?
- **Fine-tuning** the whole model **defeats the purpose** of a FM
- EO FMs **do not** (yet) **outperform supervised models**
- EO data are inherently **multi-temporal**.
  - Most approaches do not explicitly model this temporal nature and focus on single-image pretraining pipelines

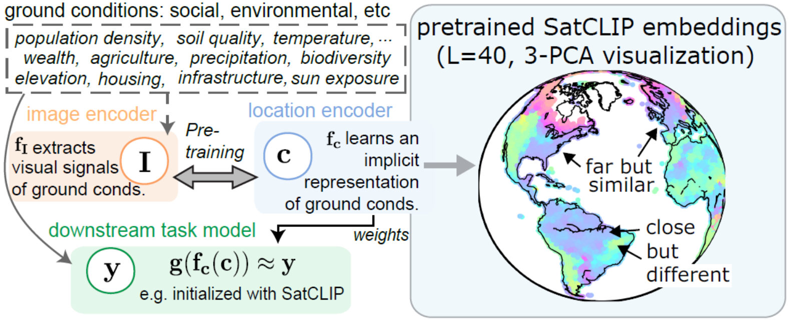
A methodological evaluation of the extracted representations is currently missing

Mechanistic interpretability at scale → understand the underlying processes of the models using interpretability techniques

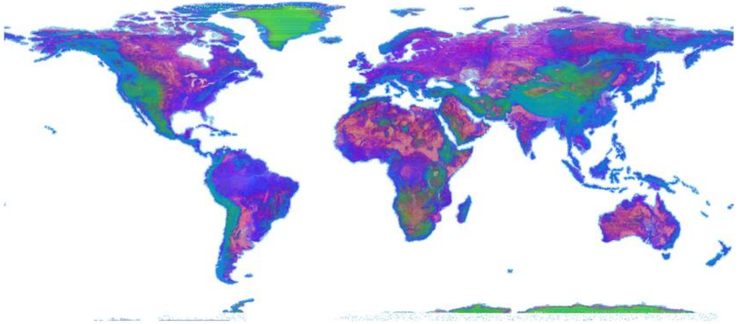
Larger does not imply better, scalability through understanding the domain better

Several unresolved challenges ahead!

Outlook: global embeddings



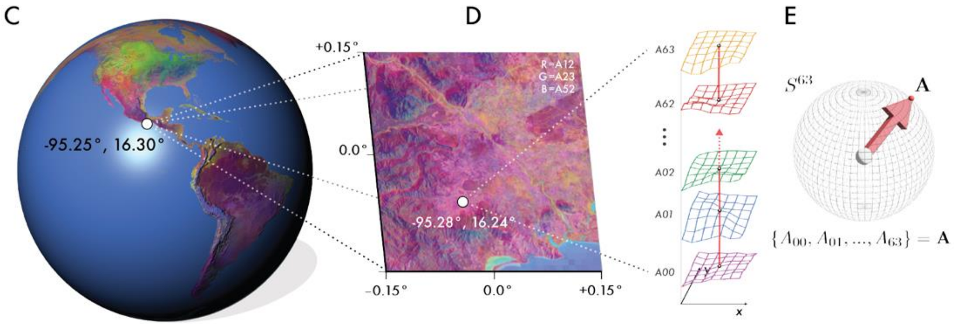
Klemmer et al. Satclip: Global, general-purpose location embeddings with satellite imagery. *AAAI* (2025)



Wang et al., Towards a unified Copernicus foundation model for earth vision. *ICCV* (2025)

"New AI model integrates petabytes of Earth observation data to generate a unified data representation that revolutionizes global mapping and monitoring"

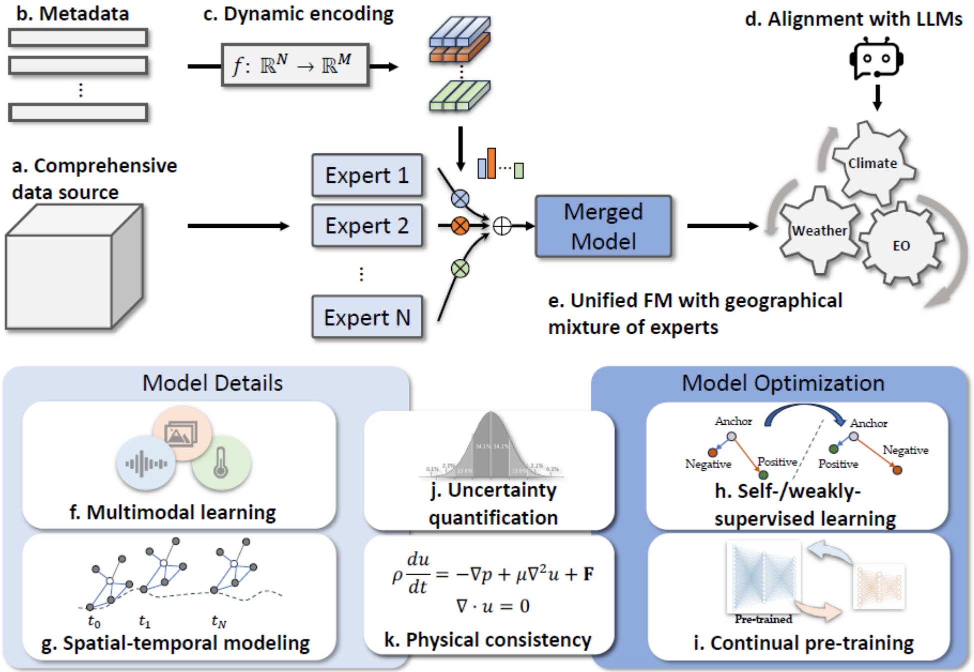
Google DeepMind



Brown et al., AlphaEarth Foundations: An embedding field model for accurate and efficient global mapping from sparse label data. arXiv (2025)

The way forward for Earth and climate FM

- Diverse Data: Satellite, reanalysis, simulations – globally and cross-modality balanced
- Metadata: Standardized for time, location, modality; enables embeddings
- Dynamic Encoding: Adapts to missing/mixed modalities using conditional computation
- LLM Alignment: Connects EO/climate models with language for reasoning & interaction.
- Geo-MoE: Geographical mixture of expert models for scalable specialization
- Multimodal Learning: Joint & modality-specific representations
- Spatio-Temporal Modeling: Scales-aware attention across time & space
- Self-/Weak Supervision: Scalable training via contrastive/predictive methods
- Continual Learning: Avoids forgetting; adapts to new data
- Uncertainty: Quantile regression, ensembles, sparse GPs
- Physics-Aware: Respects physical laws via embedded constraints



Zhu et al., On the Foundations of Earth and Climate Foundation Models. *arXiv* (2024)

Uncertainty

Probabilistic ML in Natural Hazards

Uncertainty (philosophically)

Hüllermeier et al. "Aleatoric and Epistemic Uncertainty in Machine Learning: An Introduction to Concepts and Methods." Machine Learning 110, no. 3 (March 2021): 457–506. https://doi.org/10.1007/s10994-021-05946-3.

Gawlikowski et al. "A Survey of Uncertainty in Deep Neural Networks." ArXiv:2107.03342 [Cs, Stat], July 7, 2021. http://arxiv.org/abs/2107.03342.

- **Aleatoric** (data) – notion of randomness
- **Epistemic** (model) – lack of knowledge

** alea: latin word, game of dice (random)*
episteme: greek word, meaning knowledge

Epistemic uncertainty can be reduced with more data

Aleatoric uncertainty depends on the data generation process and cannot be reduced

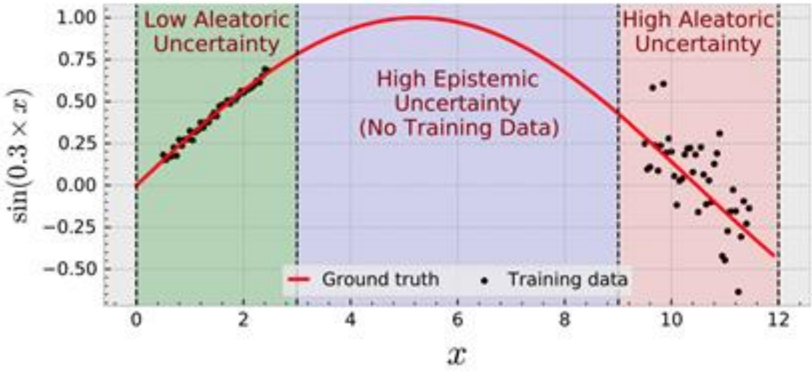
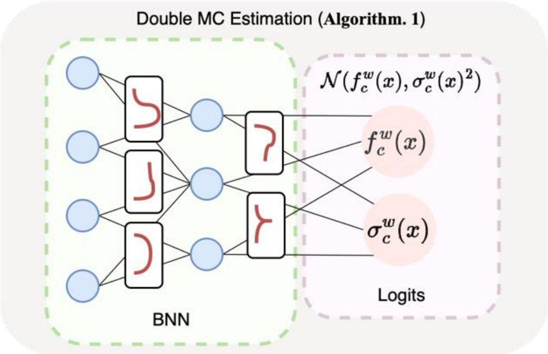


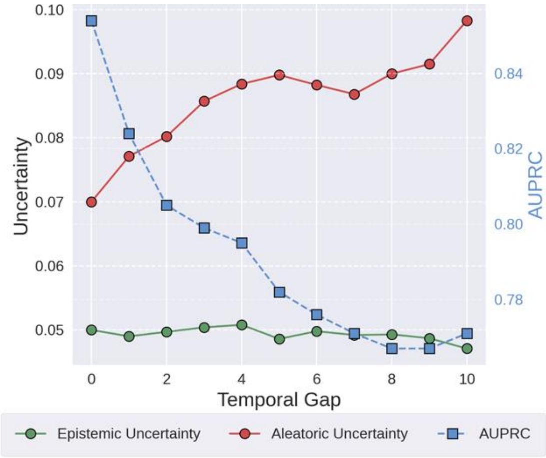
Figure 3: Illustration of the Epistemic and Aleatoric uncertainty.

Tuna et al. "Exploiting Epistemic Uncertainty of the Deep Learning Models to Generate Adversarial Samples." arXiv, February 13, 2021. https://doi.org/10.48550/arXiv.2102.04150.

An example

- Mesogeos dataset [1]
- Uncertainty-aware deep learning for wildfire danger forecasting [2]
- Capture epistemic uncertainty using Bayesian NNs
- Capture aleatoric uncertainty by accounting for the heteroscedastic label noise [3]

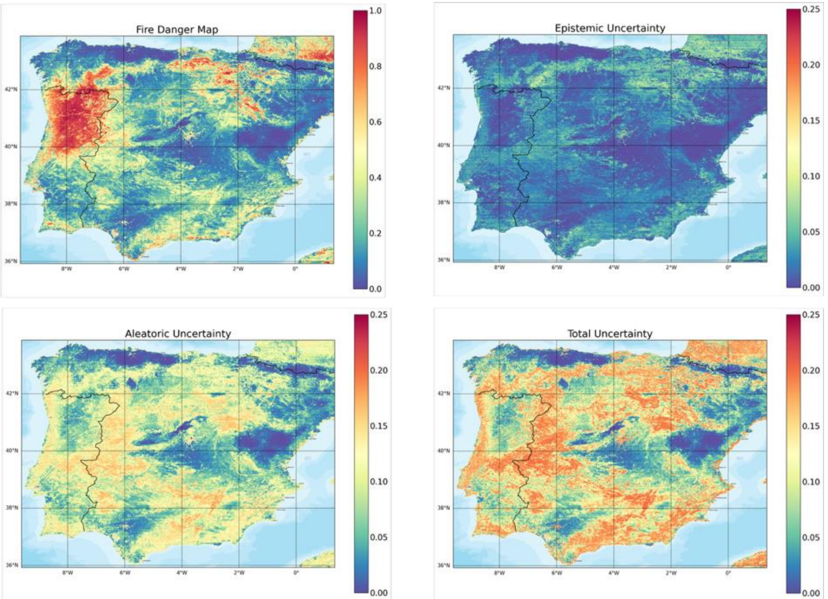




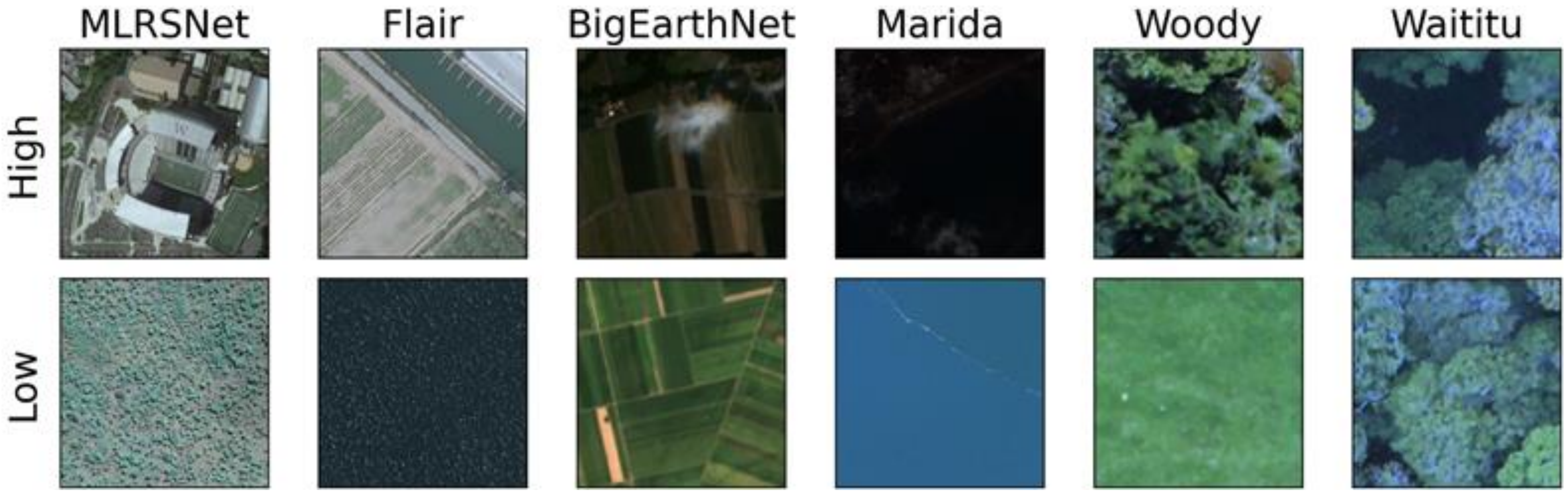
[1] Kondylatos et al., Mesogeos: A multi-purpose dataset for data-driven wildfire modeling in the mediterranean. NeurIPS (2023)

[2] Kondylatos et al., Uncertainty-aware deep learning for wildfire danger forecasting, arXiv (2025)

[3] Collier et al., A simple probabilistic method for deep classification under input-dependent label noise. *arXiv* (2020)



Zero-shot uncertainty estimation



Foundation models for Earth Observation

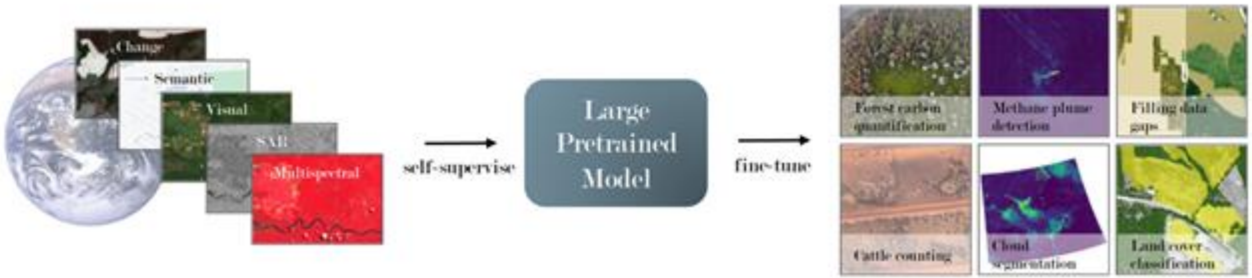


Image taken from [1].

Lacoste et al., Geo-bench: Toward foundation models for earth monitoring. *Advances in NeurIPS* (2023)

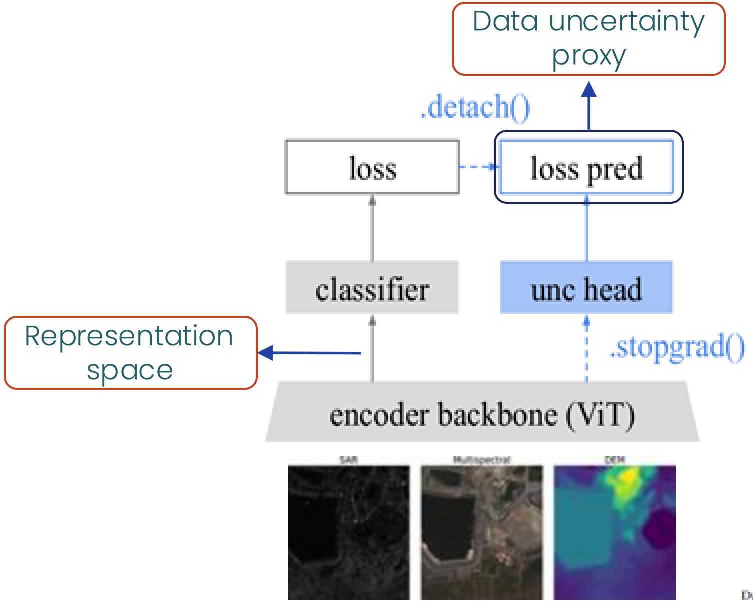
Generalizable uncertainty estimations

Image adapted from [1].

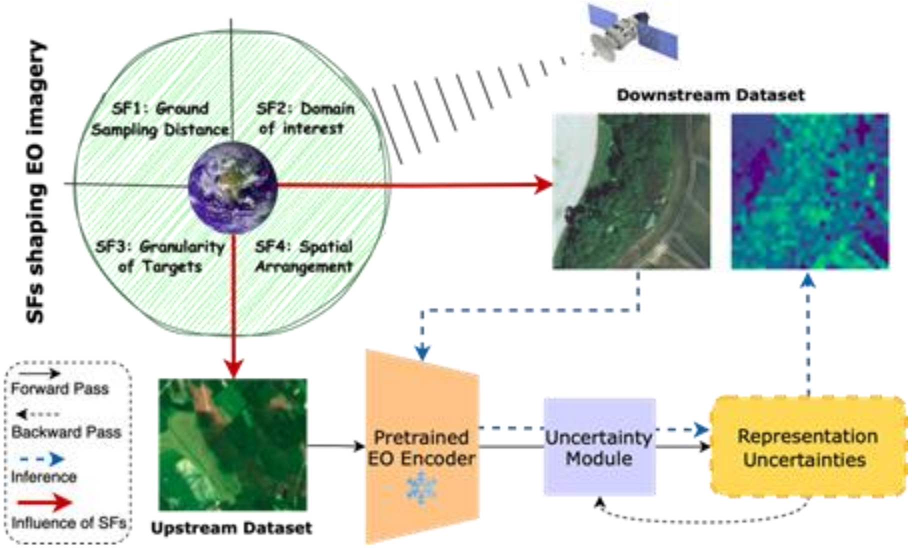
[1] Lacoste et al., Geo-bench: Toward foundation models for earth monitoring. *Advances in NeurIPS* (2023)

Pretrained representation uncertainties

Kondylatos S., Bountos N.I. et al., On the Generalization of Representation Uncertainty in Earth Observation. ICCV (2025)



Kirchhof et al., Pretrained visual uncertainties. *arXiv* (2024)



Dataset	Input Modality	Sensor	ML Setup	#Classes	EO task	Spatial Res.	Pretraining	Inference	Image Size	Coverage
Imagenet	RGB	Optical	Classification	21, 841	Optical Images	-	✔	✘	224 × 224	-
BigEarthNet	MS/SAR	S1, S2	Multi-label classification	19	LULC classification	10m	✔	✔	120 × 120	Europe
BigEarthNet-5	MS/SAR	S1, S2	Multi-label classification	5	LULC classification	10m	✔	✘	120 × 120	Europe
MLRSNet	RGB	Multi-sensor	Multi-label classification	60	Semantic Scene Understanding	≈10 - 0.1m	✘	✔	256 × 256	Global
Woody	RGB	Aerial	Image Segmentation	4	Tree-species detection	50cm	✘	✔	224 × 224	Chile
Waititu	RGB	Aerial	Image Segmentation	3	Invasion tree-species detection	50cm	✘	✔	224 × 224	New Zealand
Flair	RGB/NIR/DEM	Aerial	Image Segmentation	19	LULC classification	20cm	✔	✔	512 × 512	France
Marida	MS	S2	Image Segmentation	12	Marine Debris Detection	10m	✘	✔	224 × 224	Global

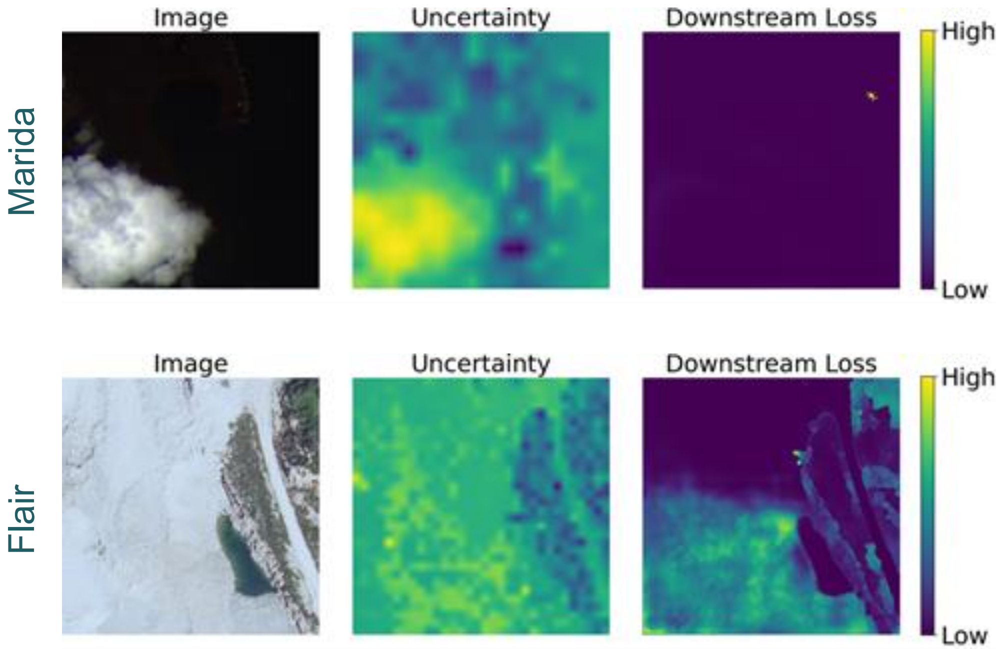
Zero shot uncertainty estimation

Kondylatos S., Bountos N.I. et al., On the Generalization of Representation Uncertainty in Earth Observation. ICCV (2025)

Can we go beyond sample-based uncertainty estimations?

Kondylatos S., Bountos N.I. et al., On the Generalization of Representation Uncertainty in Earth Observation. ICCV (2025)

- Each ViT patch can be considered as an EO image
- Estimating uncertainties for each patch results in localized uncertainty estimation



Flair and Marida uncertainties estimated from BigEarthNet pretrained ViT-Large vis-a-vis downstream pixel loss.

More about us…

- https://orion-ai-lab.github.io/
- https://github.com/orion-ai-lab
- Spring School: AI for Modeling and Understanding Climate Extremes
- ThiningEarth

Thank you for your attention

Ioannis Papoutsis

ipapoutsis@mail.ntua.gr