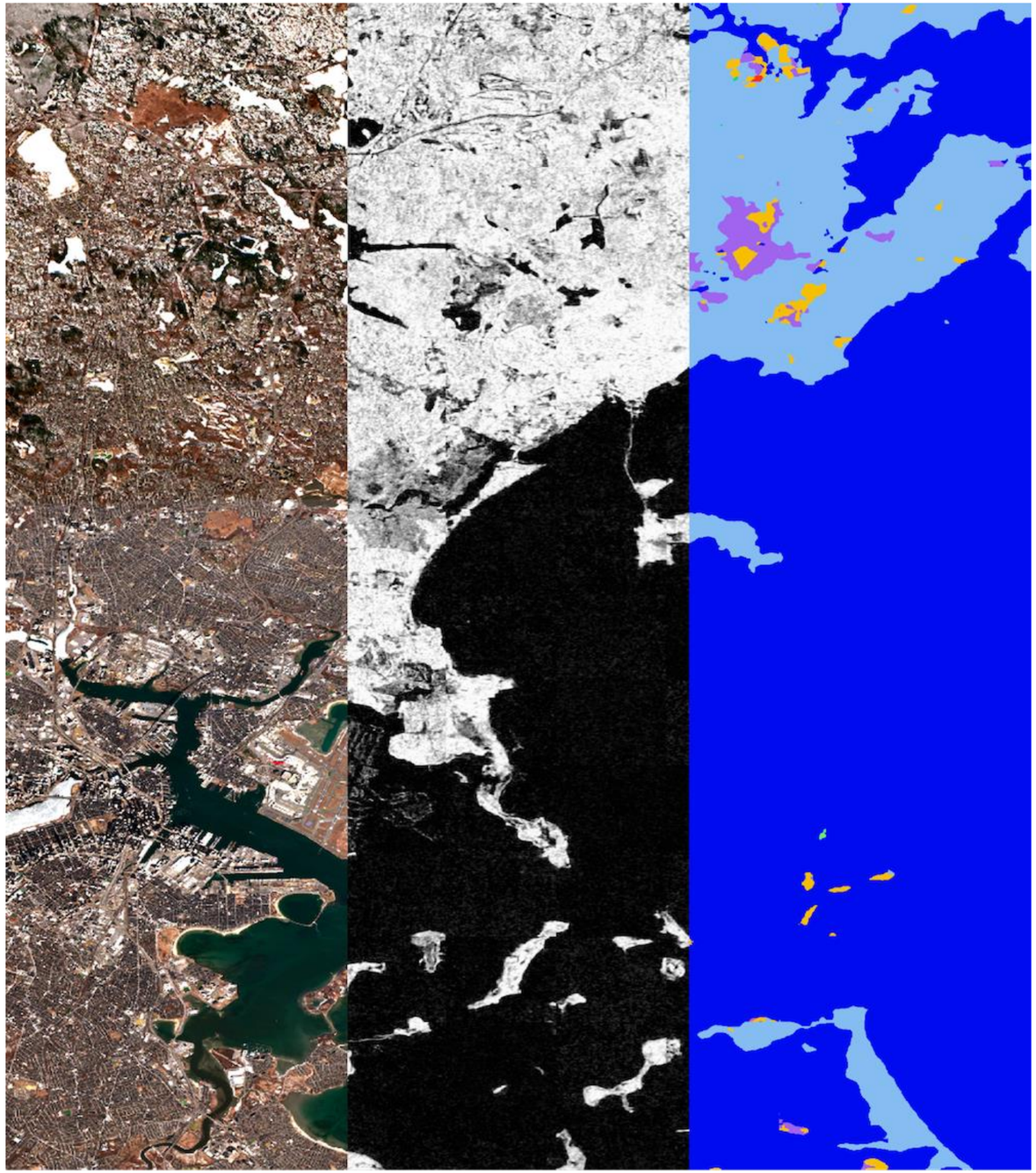
TerraMind



Dr. Johannes Jakubik

IBM

esa Φ-lab

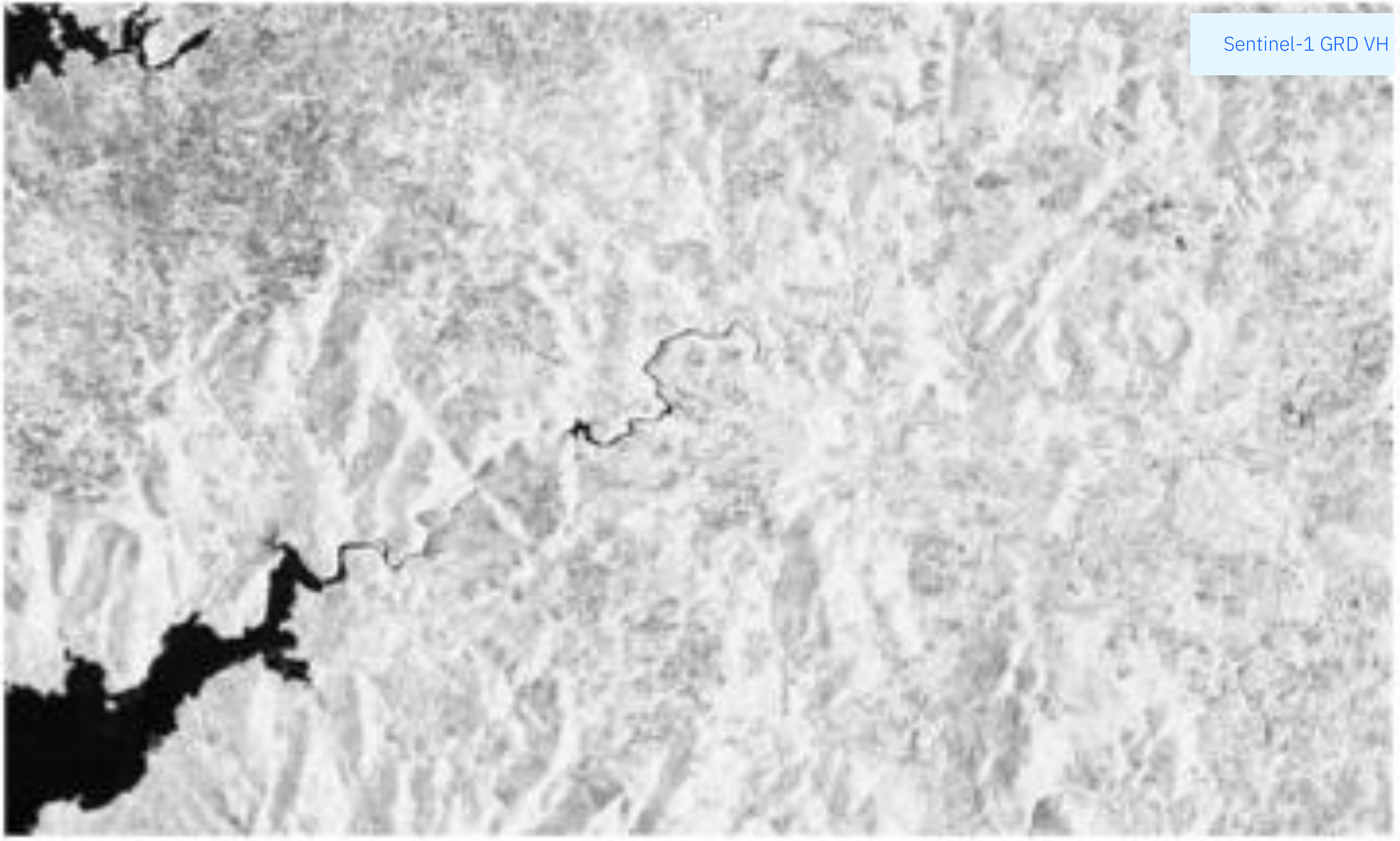
Turing test

Input:
Optical data

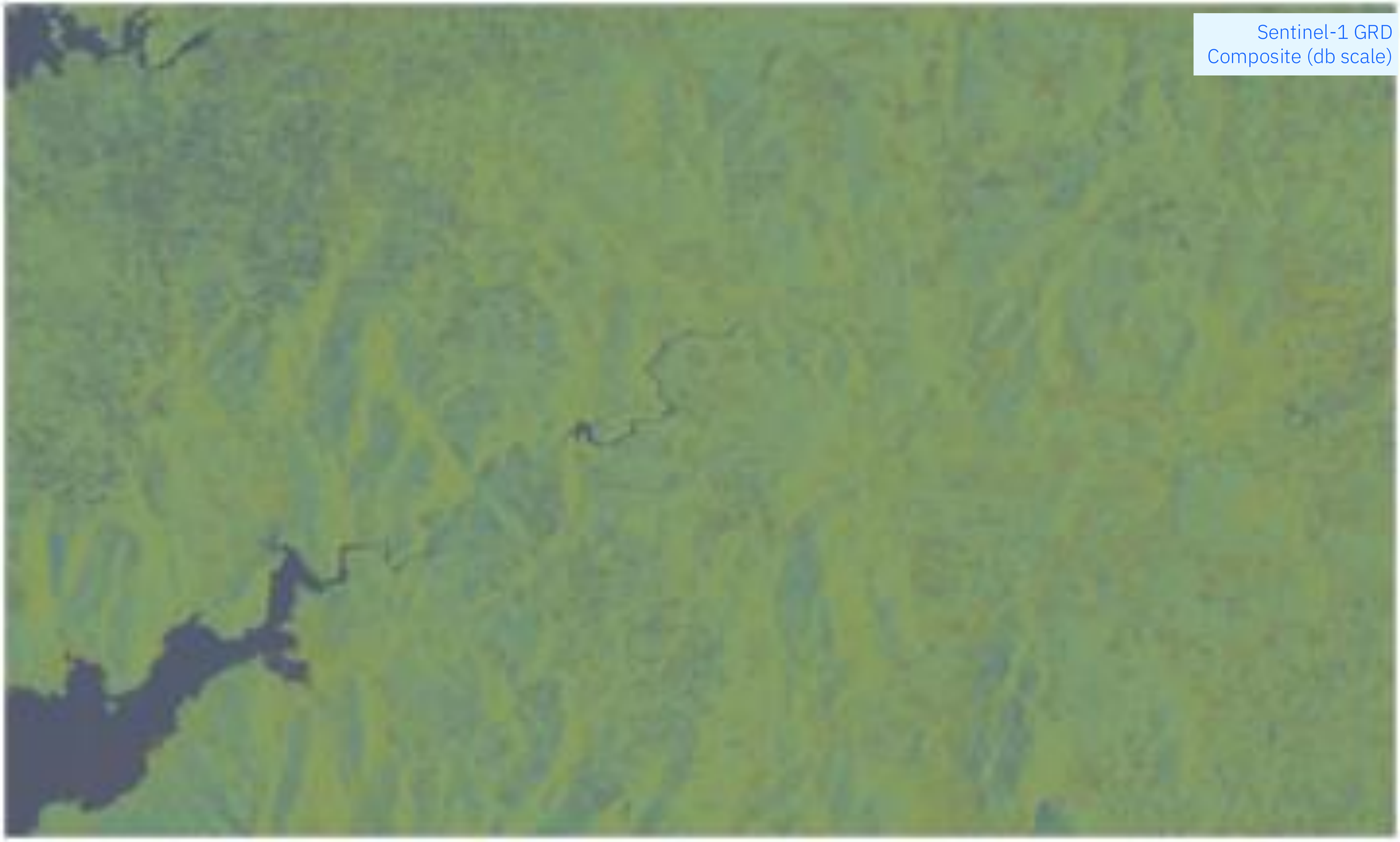

Sentinel-2 L2A

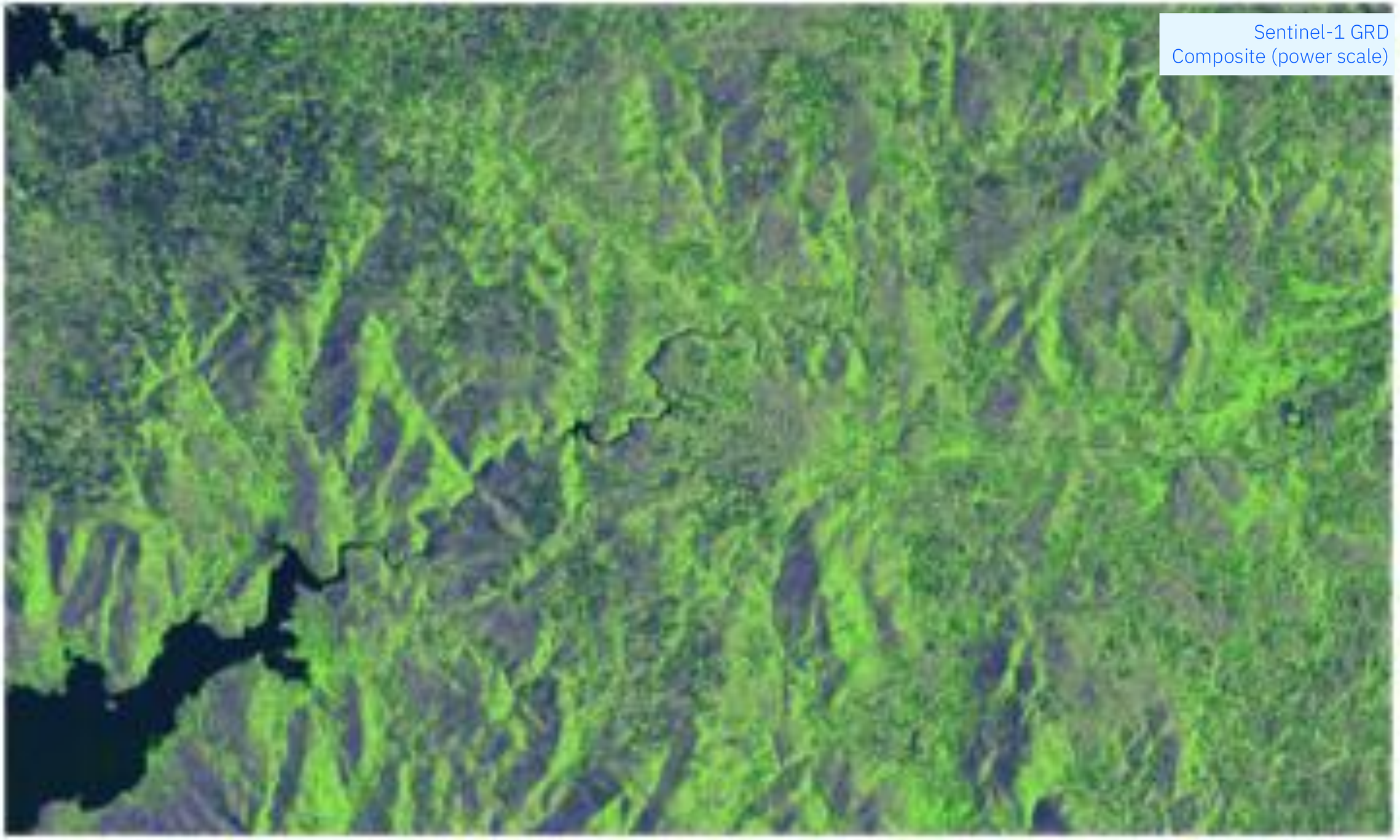
Real sensor data or generated?

S1RTC

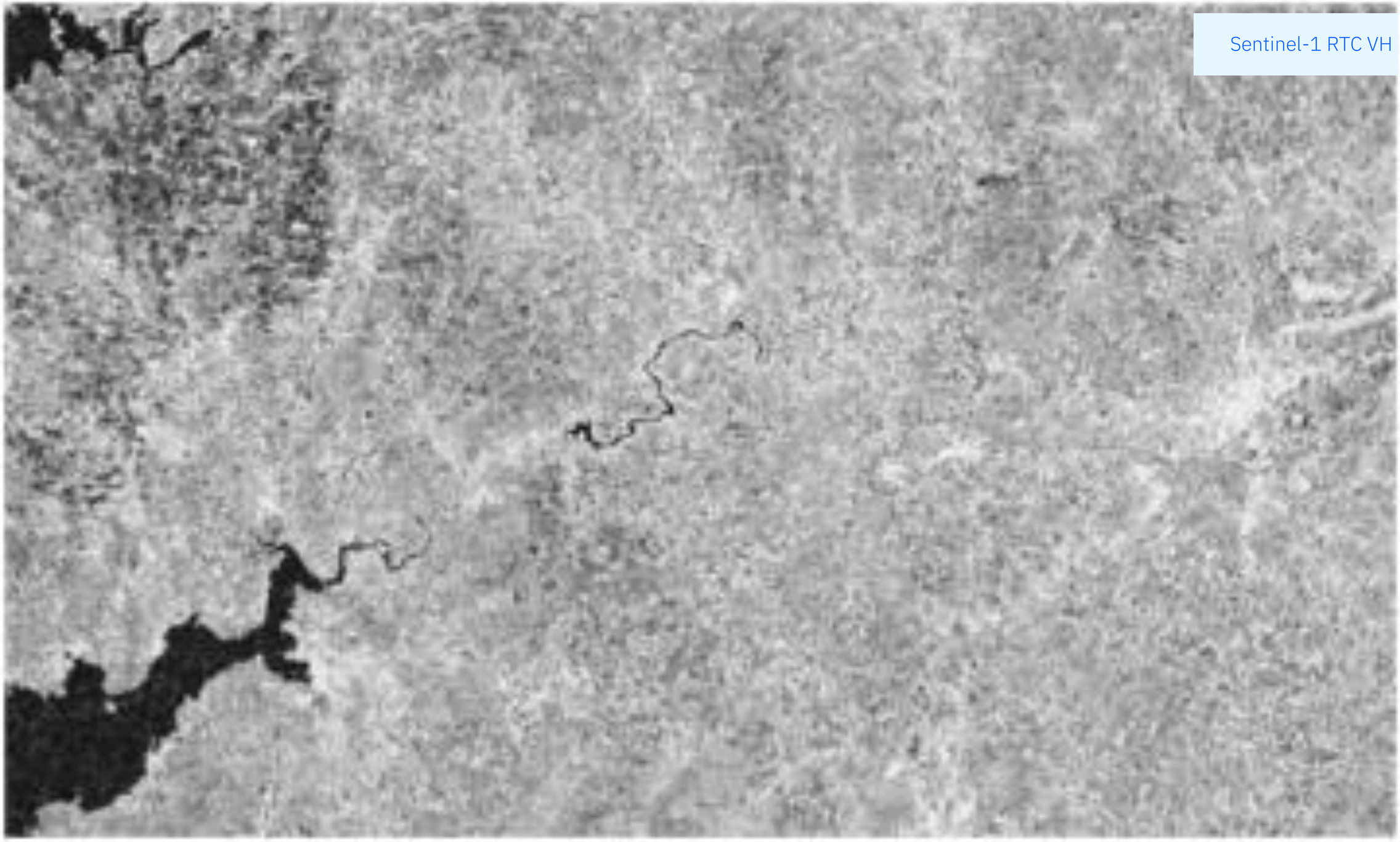

Sentinel-1 GRD VH


Sentinel-1 GRD VV


Sentinel-1 GRD
Composite (db scale)

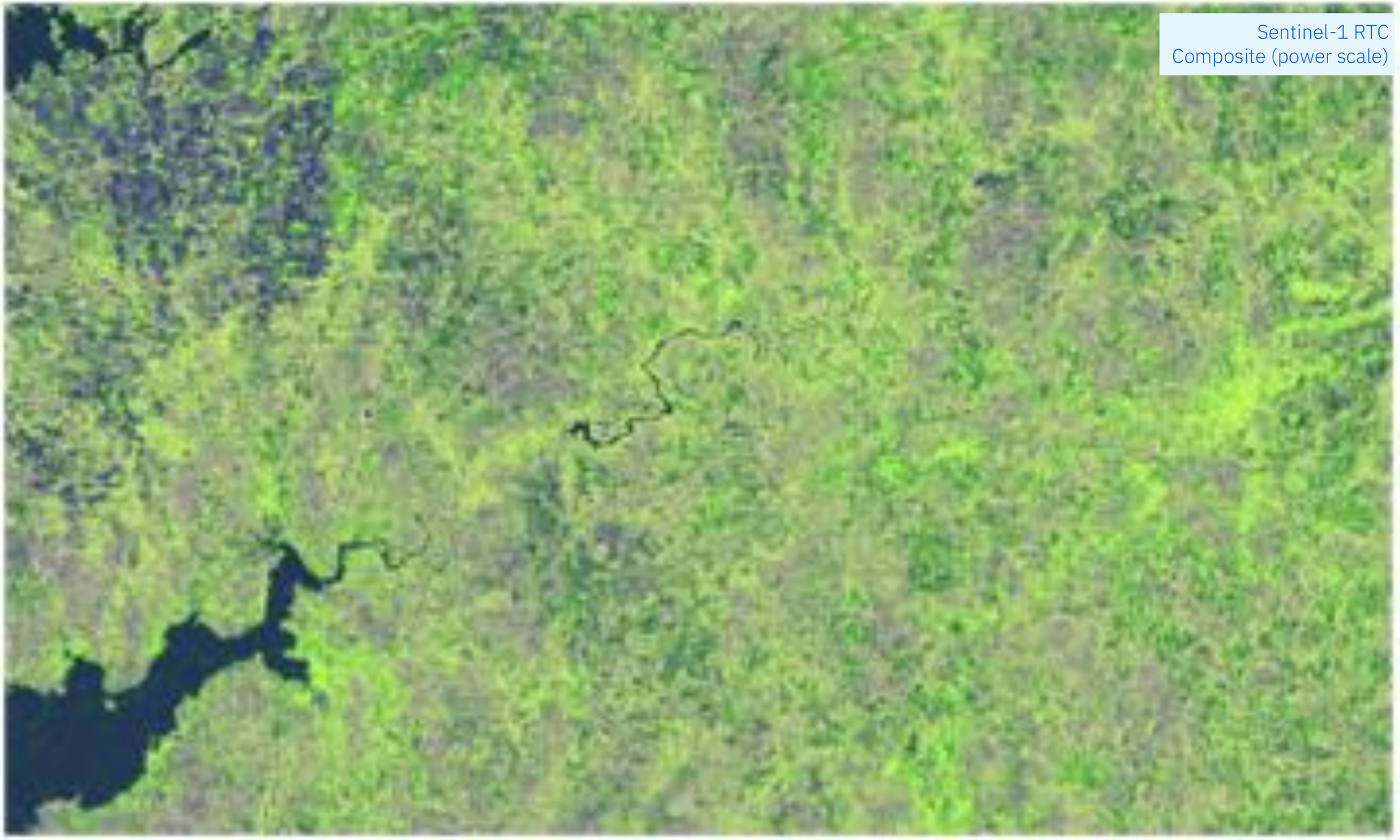

Sentinel-1 GRD
Composite (power scale)

S1GRD

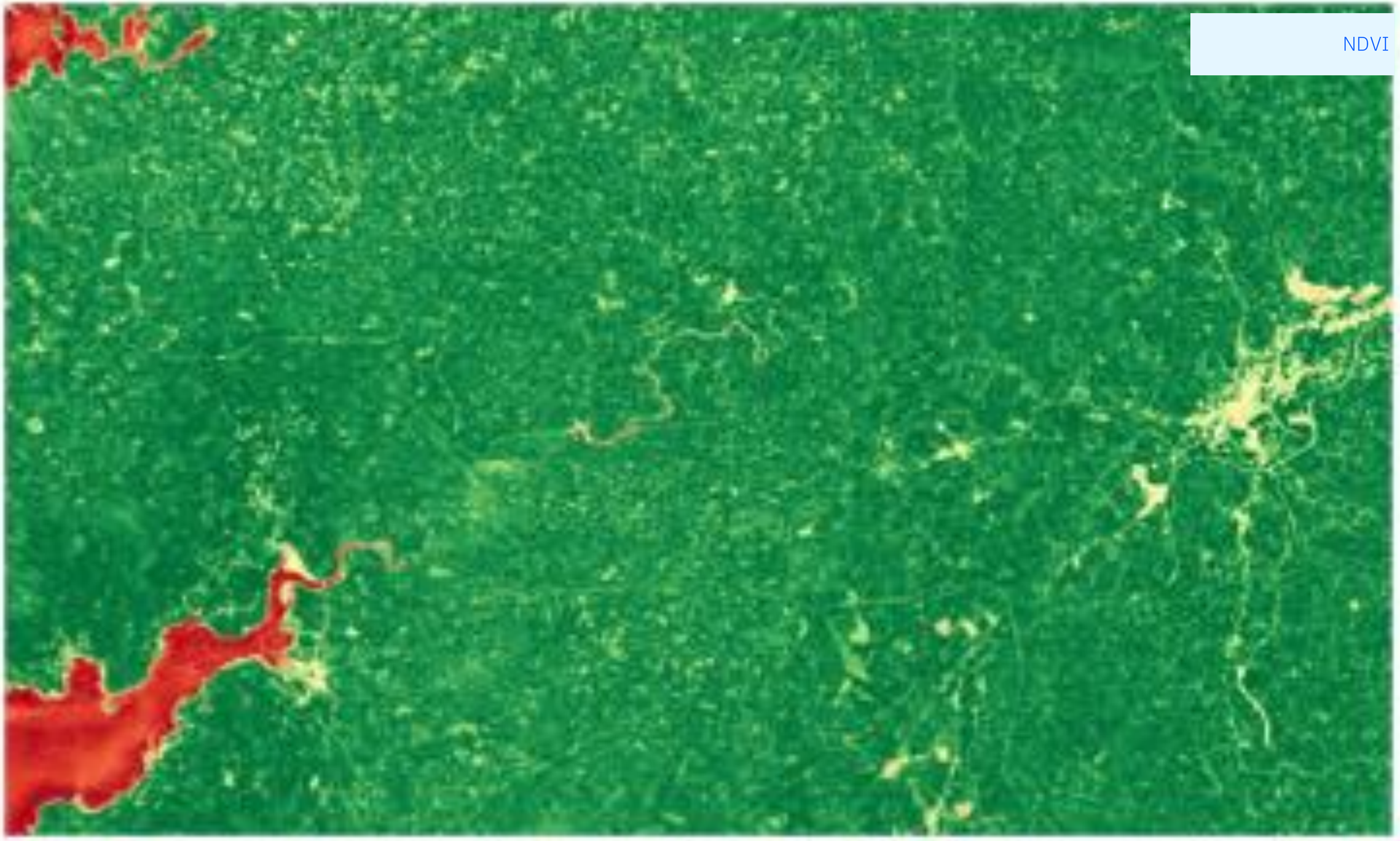

Sentinel-1 RTC VH


Sentinel-1 RTC VV

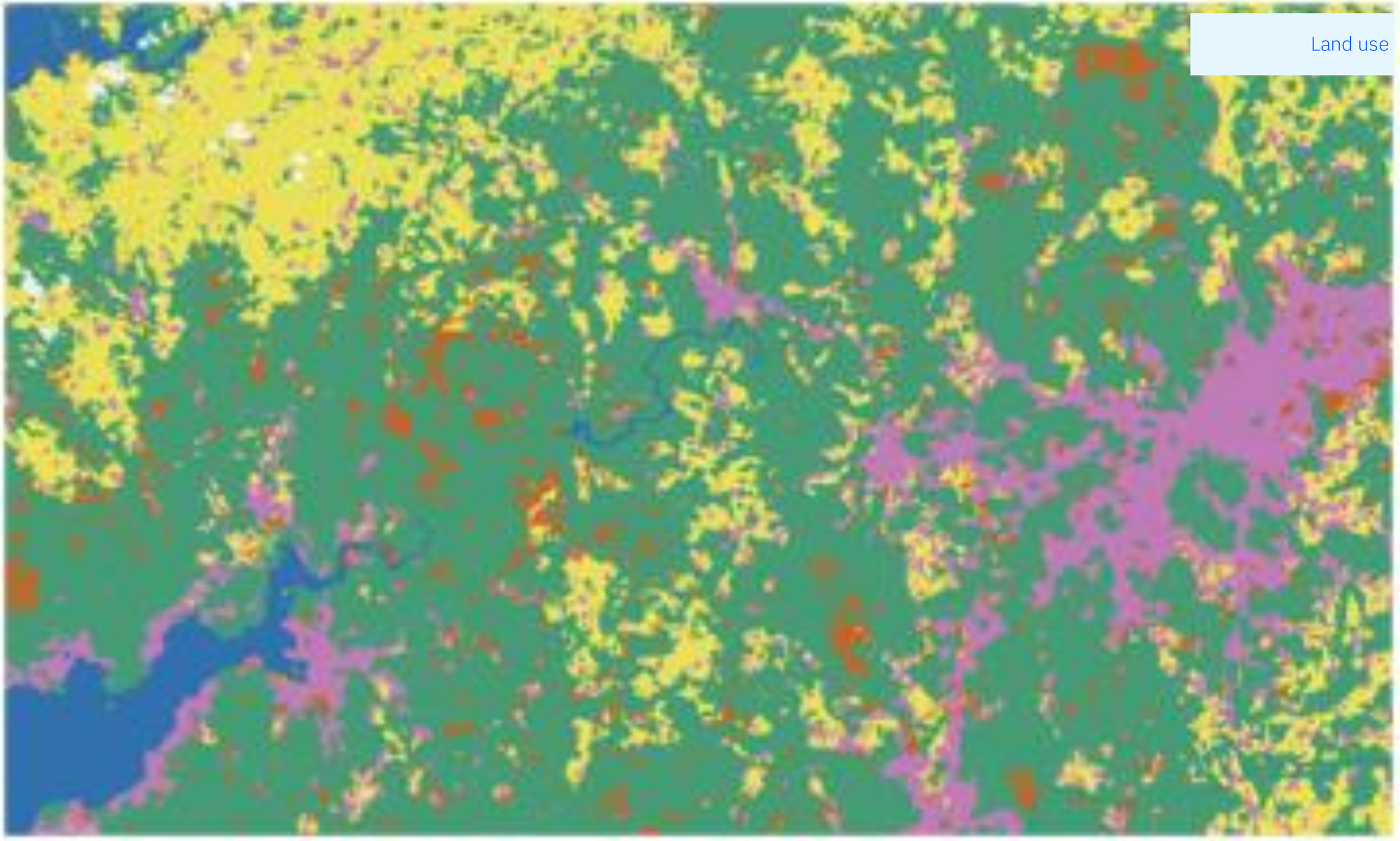

Sentinel-1 RTC
Composite (db scale)


Sentinel-1 RTC
Composite (power scale)

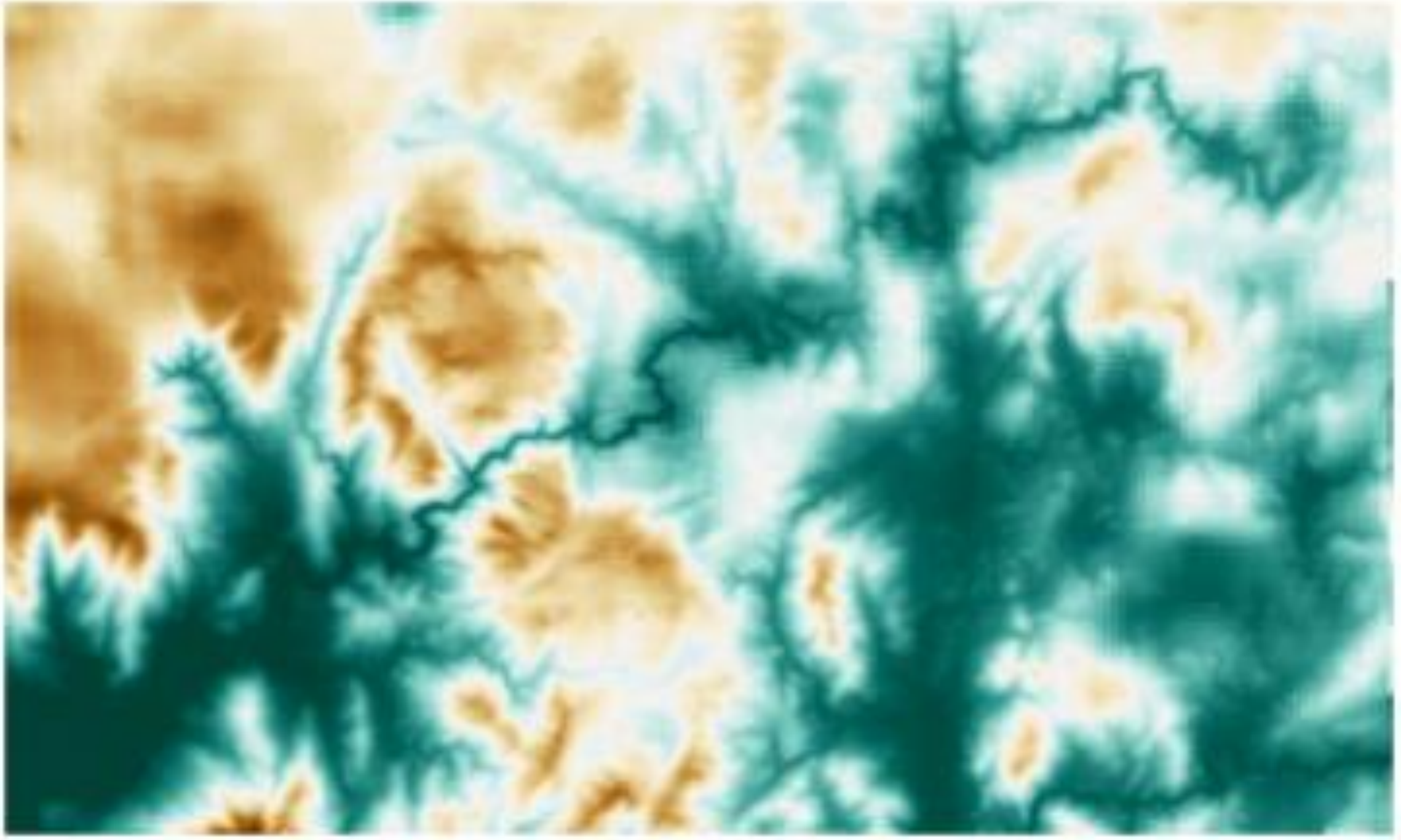
NDVI


NDVI

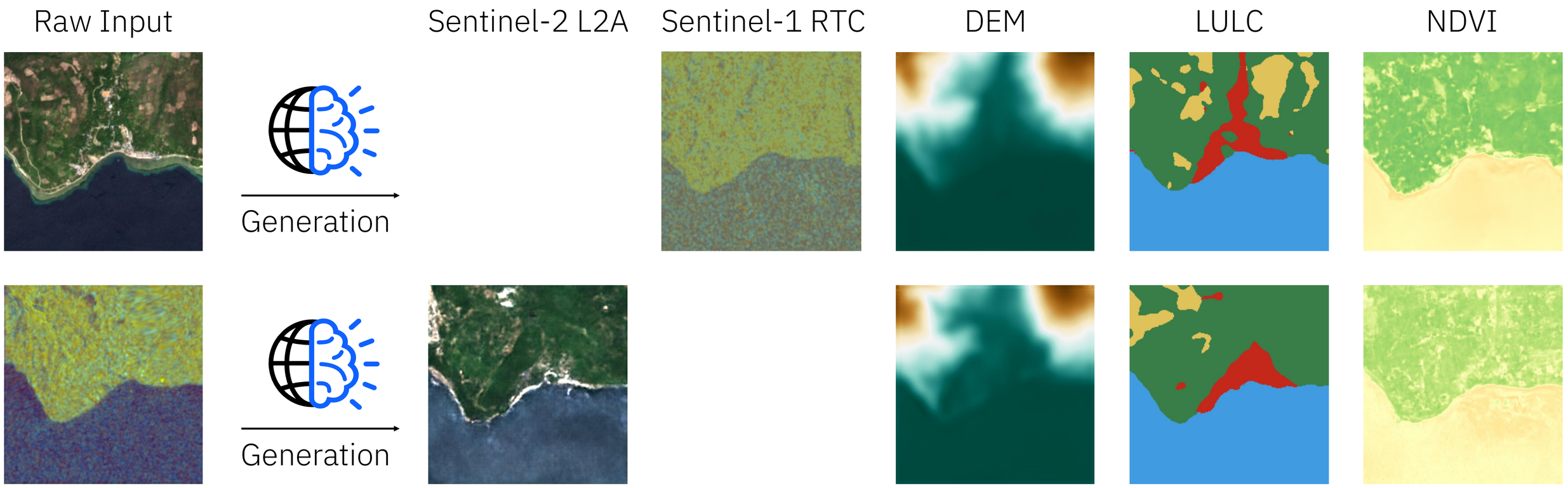
Land use


Land use

Digital elevation

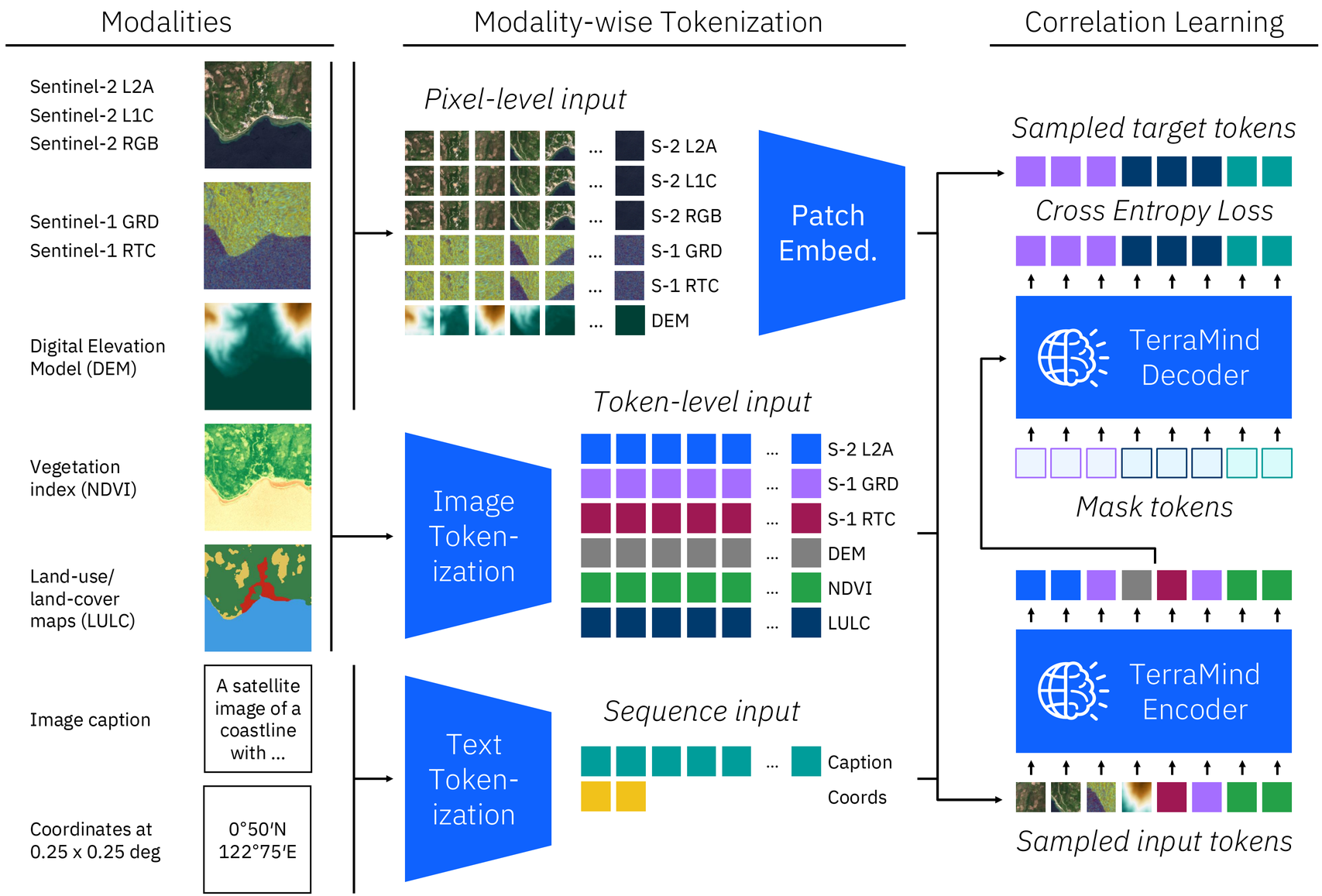


TerraMind – our first foundation model with cross-modal understanding



TerraMind

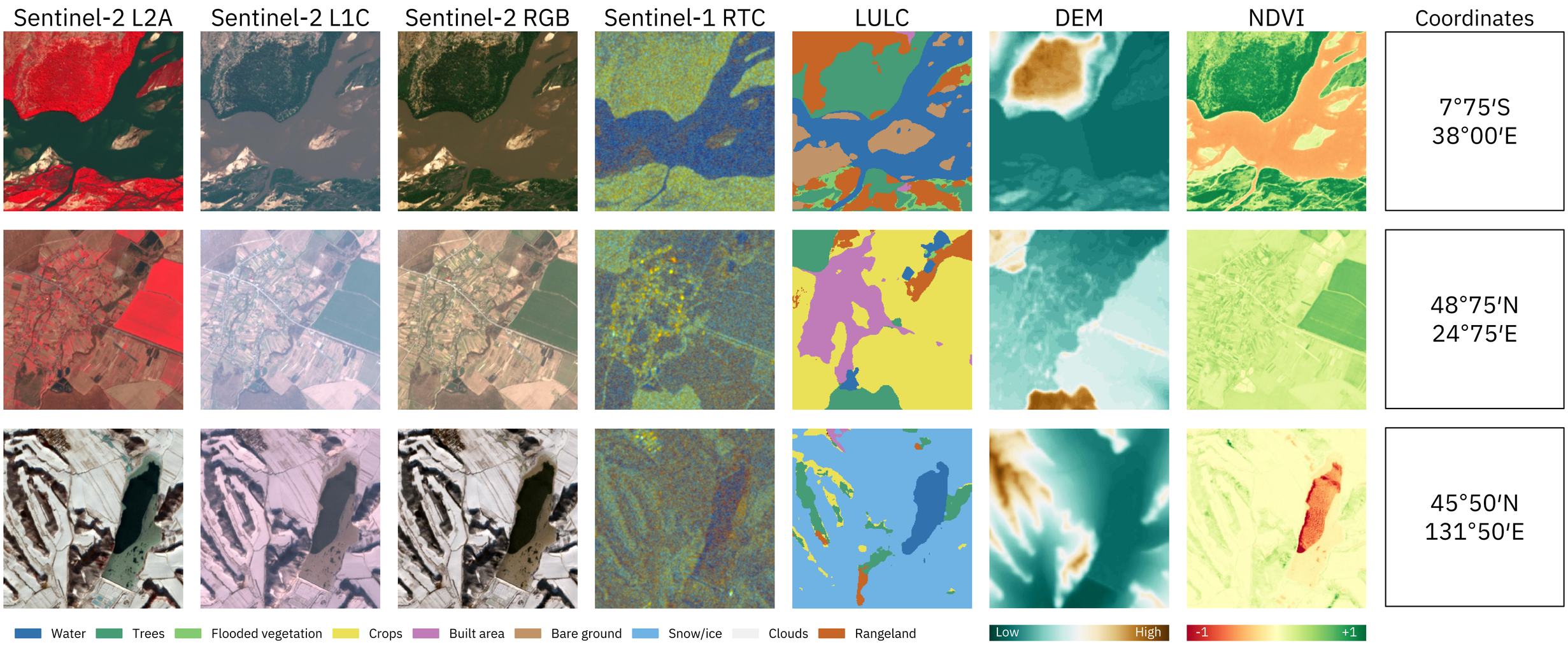
TerraMind is our first any-to-any generative, large-scale multimodal FM for EO and is pre-trained on 500 billion tokens using diverse geospatial data.



Jakubik, J., Yang, F., Blumenstiel, B., Scheurer, E., Sedona, R., Maurogiovanni, S., ... & Longépé, N. (2025). TerraMind: Large-Scale Generative Multimodality for Earth Observation. *arXiv preprint arXiv:2504.11171*.

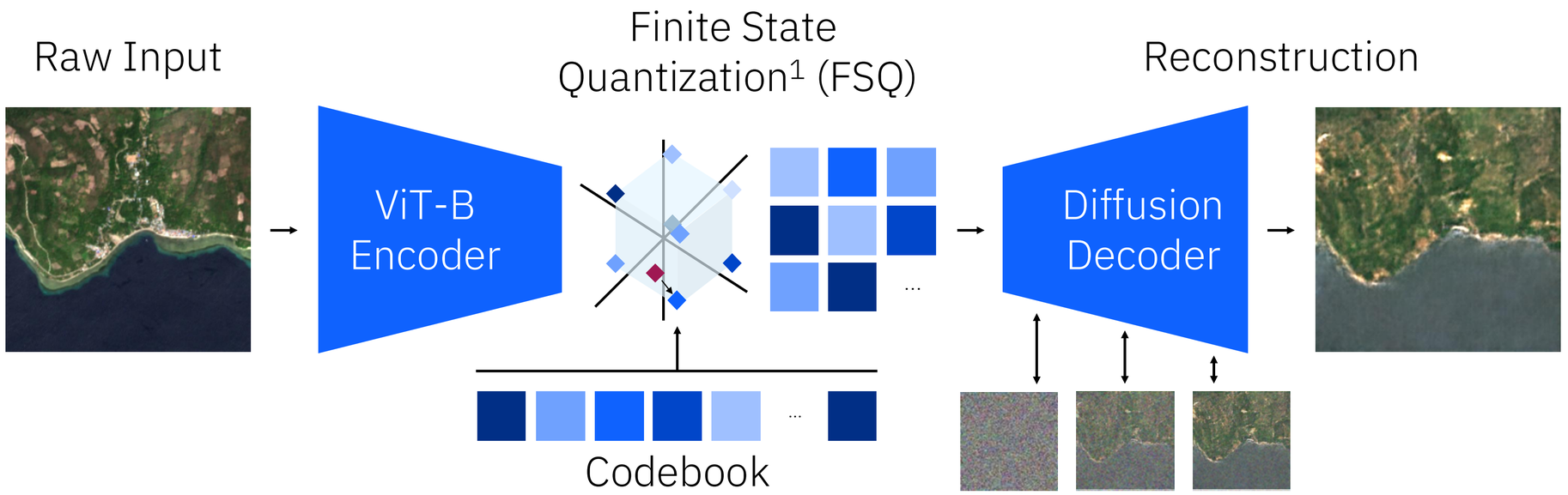
The training data – 9M spatio-temporal locations with multiple modalities from TerraMesh dataset

Open sourcing in June



Blumenstiel, B., Fraccaro, P., Marsocci, V., Jakubik, J., Maurogiovanni, S., Czerkawski, M., ... & Longépé, N. (2025). TerraMesh: A Planetary Mosaic of Multimodal Earth Observation Data. arXiv preprint arXiv:2504.11172.

Tokens are good pre-training targets

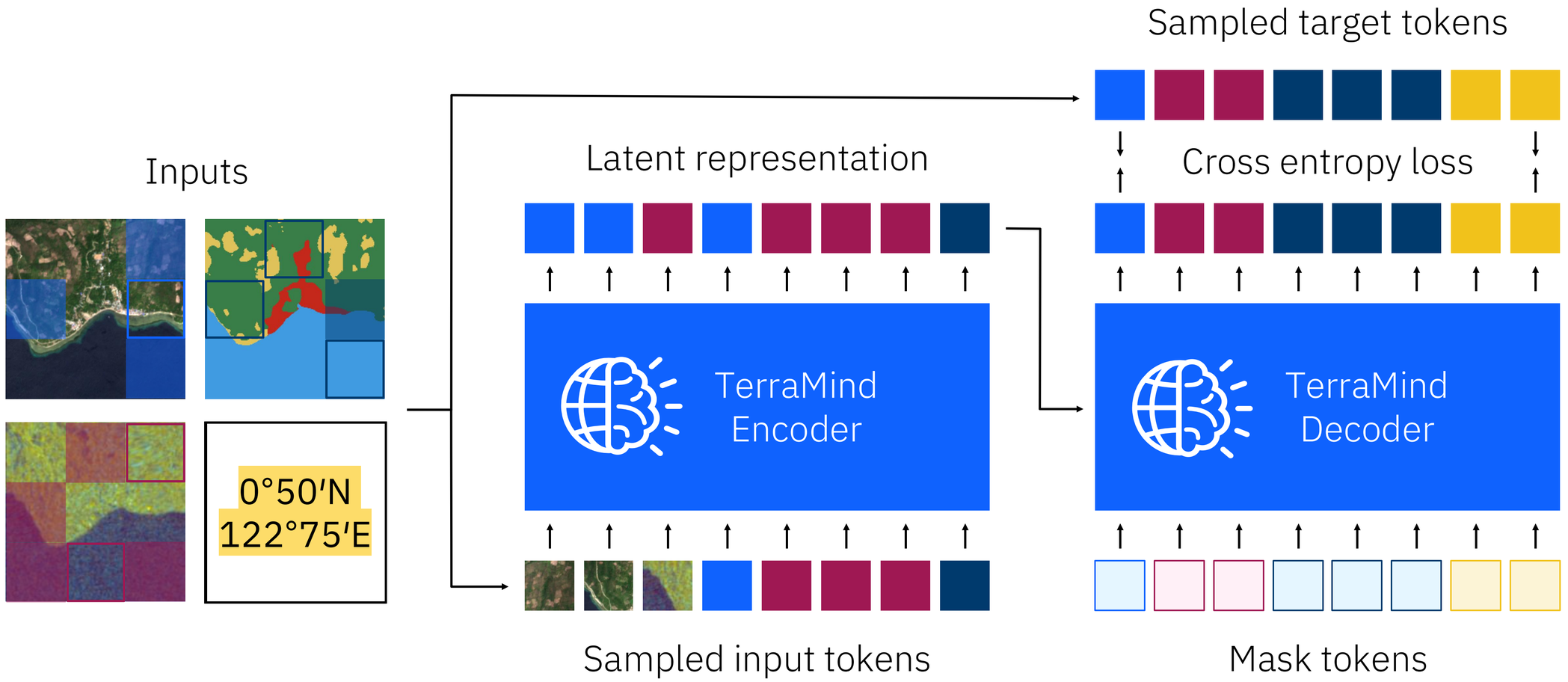


TerraMind uses Vector-quantized Variational Auto-Encoders (VQ-VAE) with diffusion steps for tokenization.

The tokenizer of each modality is trained separate and uses Finite Sate Quantization (FSQ) with a codebook size of 15,360 tokens.

1 Mentzer, F., Minnen, D., Agustsson, E., & Tschannen, M. (2023). Finite scalar quantization: Vq-vae made simple. *arXiv preprint arXiv:2309.15505*.

Fusing the modalities with correlation learning

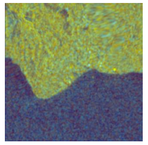


Mizrahi, D., Bachmann, R., Kar, O., Yeo, T., Gao, M., Dehghan, A., & Zamir, A. (2023). 4m: Massively multimodal masked modeling. *Advances in Neural Information Processing Systems*.

Modalities

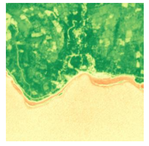
Sentinel-2 L2A
Sentinel-2 L1C
Sentinel-2 RGB

Sentinel-1 GRD
Sentinel-1 RTC



Digital Elevation Model (DEM)

Vegetation index (NDVI)



Land-use/ land-cover maps (LULC)

Image caption

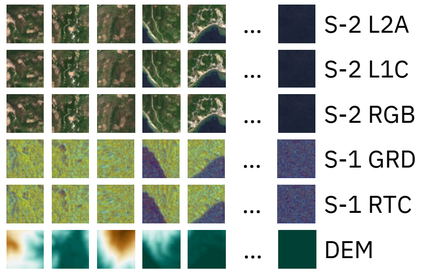
A satellite image of a coastline with ...

Coordinates at 0.25 x 0.25 deg

0°50′N 122°75′E

Modality-wise Tokenization

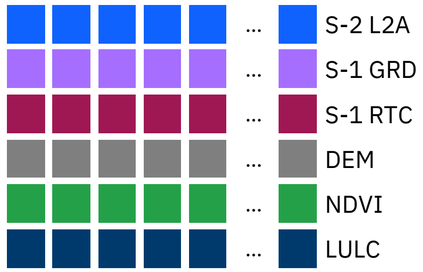
Pixel-level input



Patch Embed.

Token-level input

Image Token-ization



Sequence input

Text Token-ization

Correlation Learning

Sampled target tokens

Cross Entropy Loss

TerraMind Decoder

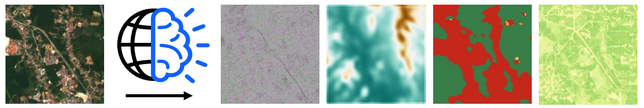
Mask tokens

TerraMind Encoder

Sampled input tokens

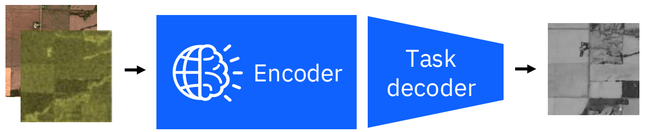
Downstream Applications

Multimodal Generation



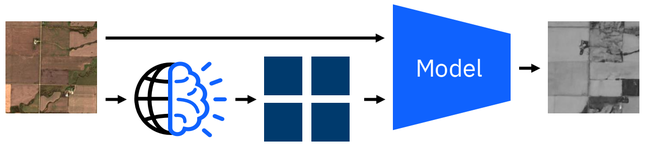
Generate data based on learned correlations

Fine-tuning



Native support for multi-modal inputs

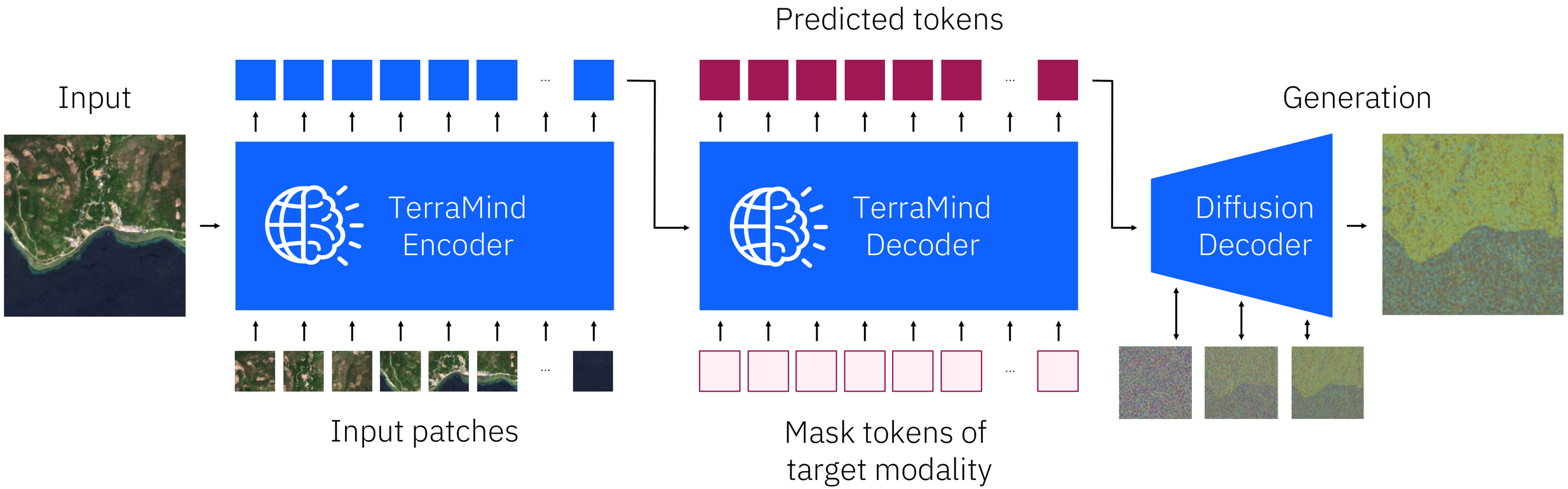
Thinking in Modalitites (TiM)



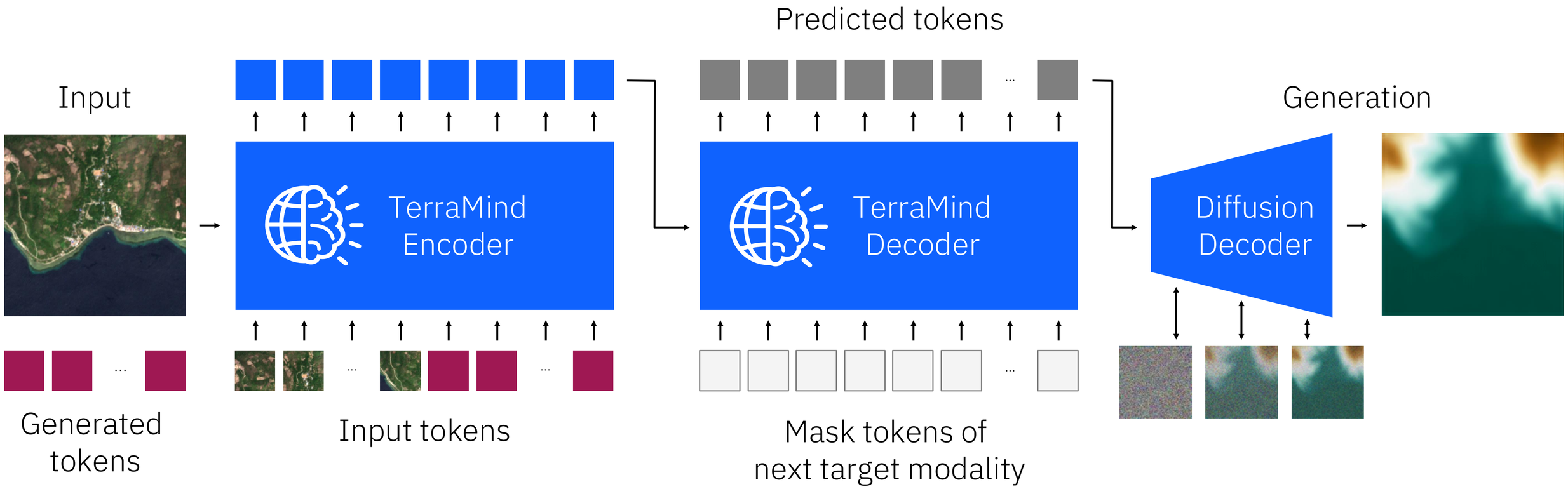
TiM tuning with generated modalities

Jakubik, J., Yang, F., Blumenstiel, B., Scheurer, E., Sedona, R., Maurogiovanni, S., ... & Longépé, N. (2025). TerraMind: Large-Scale Generative Multimodality for Earth Observation. *arXiv preprint arXiv:2504.11171*.

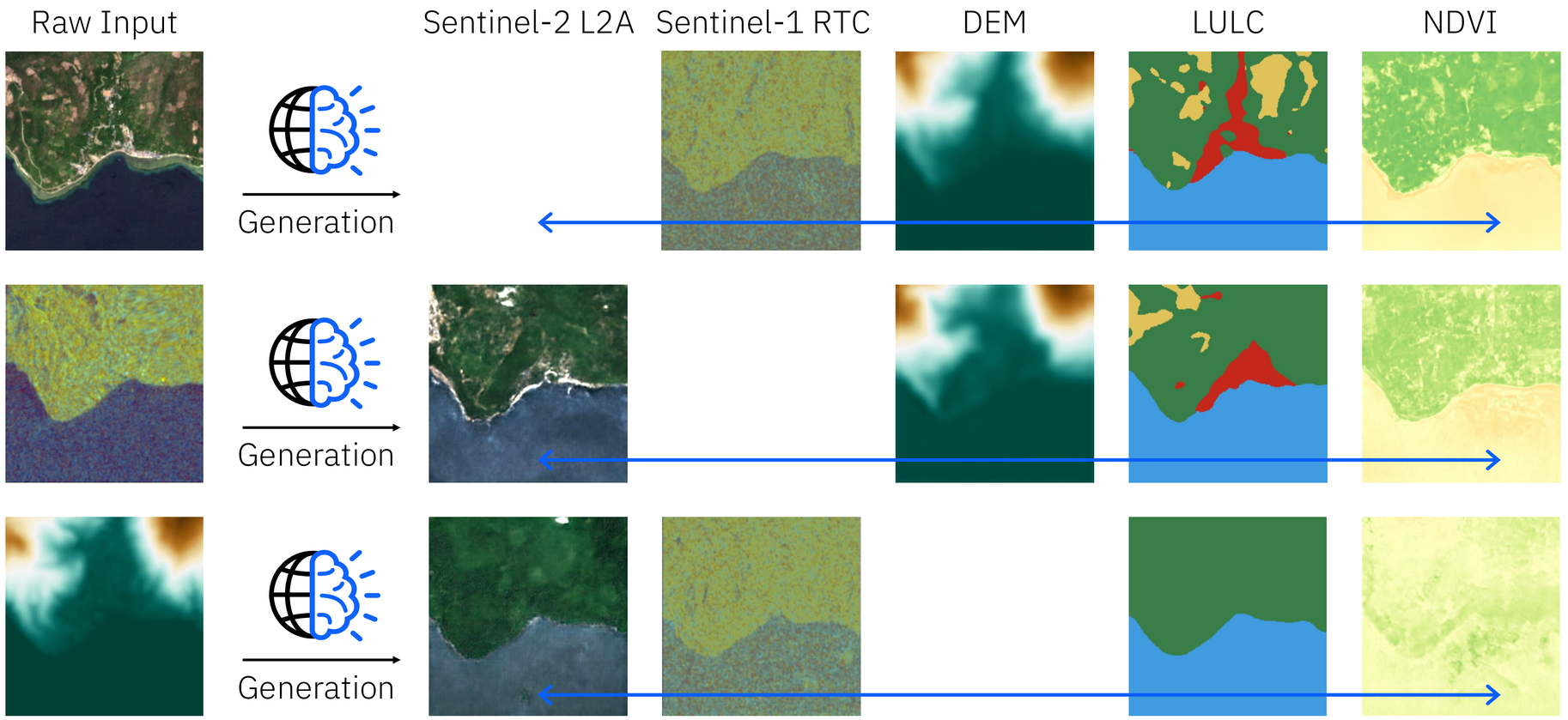
Any-to-any generation



Chained generation for consistent generations



Chained generation for consistent generations



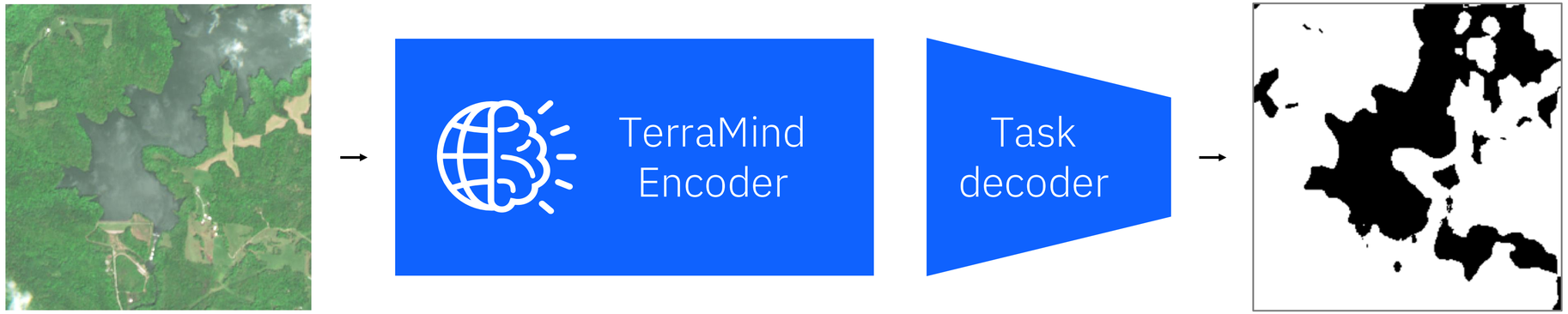
Consistancy between generated modalities

Thinking in Modalities

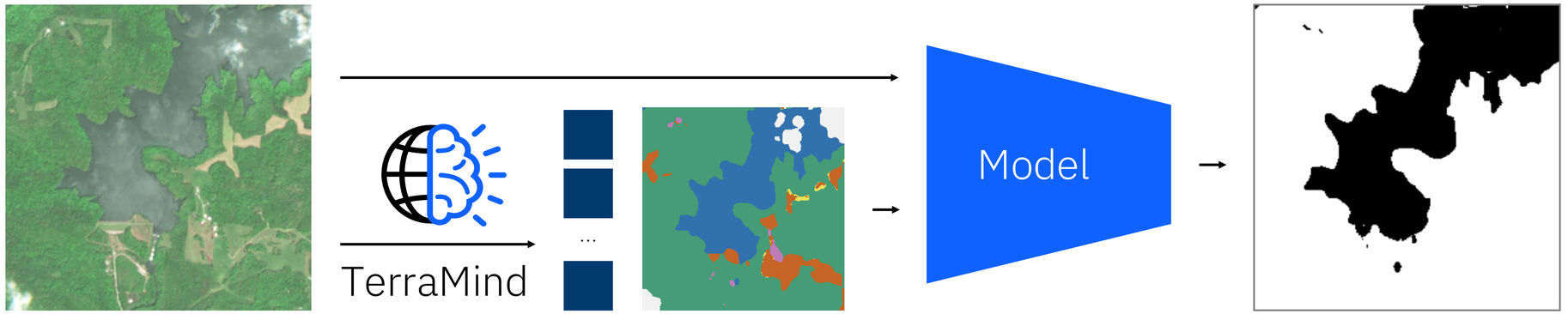
TerraMind enables to enhance fine-tuning by Thinking-in-Modalities (TiM) – generating intermediate artificial data of other modalities.

The raw image and the generated tokens are used as input by the fine-tuned model.

Standard fine-tuning

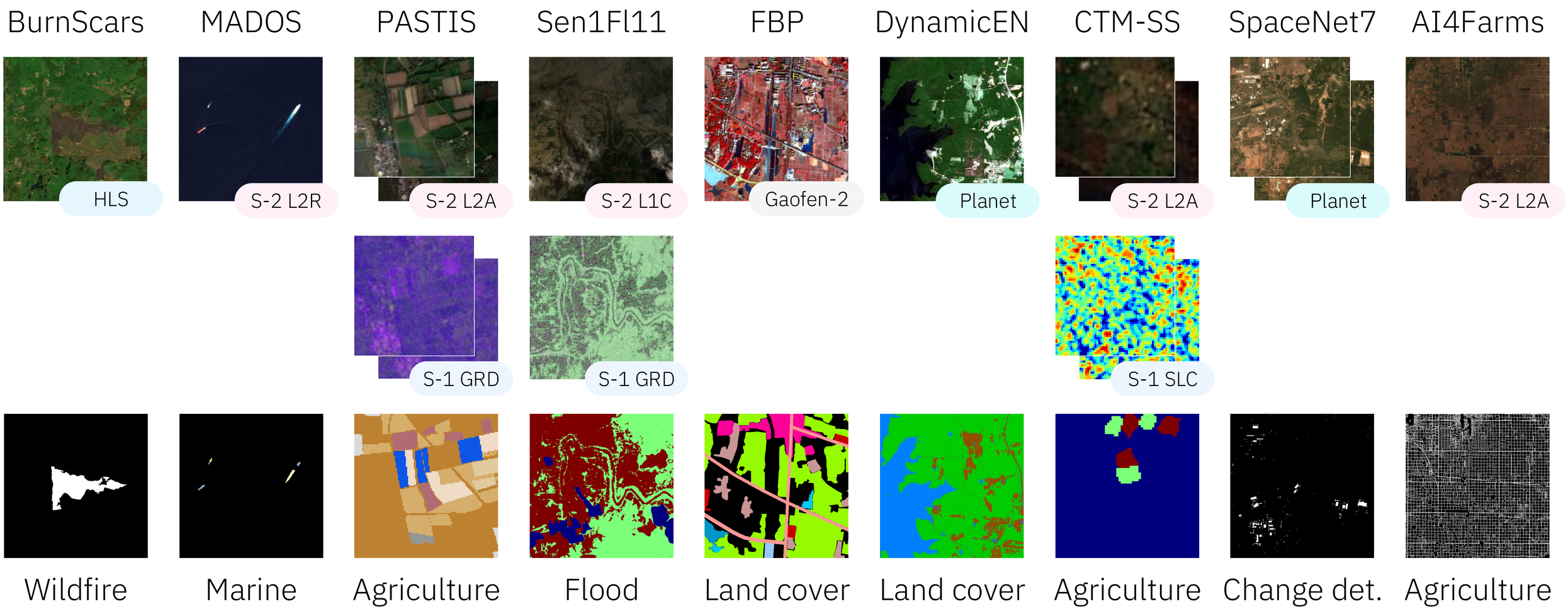


TiM fine-tuning with intermediate modalities



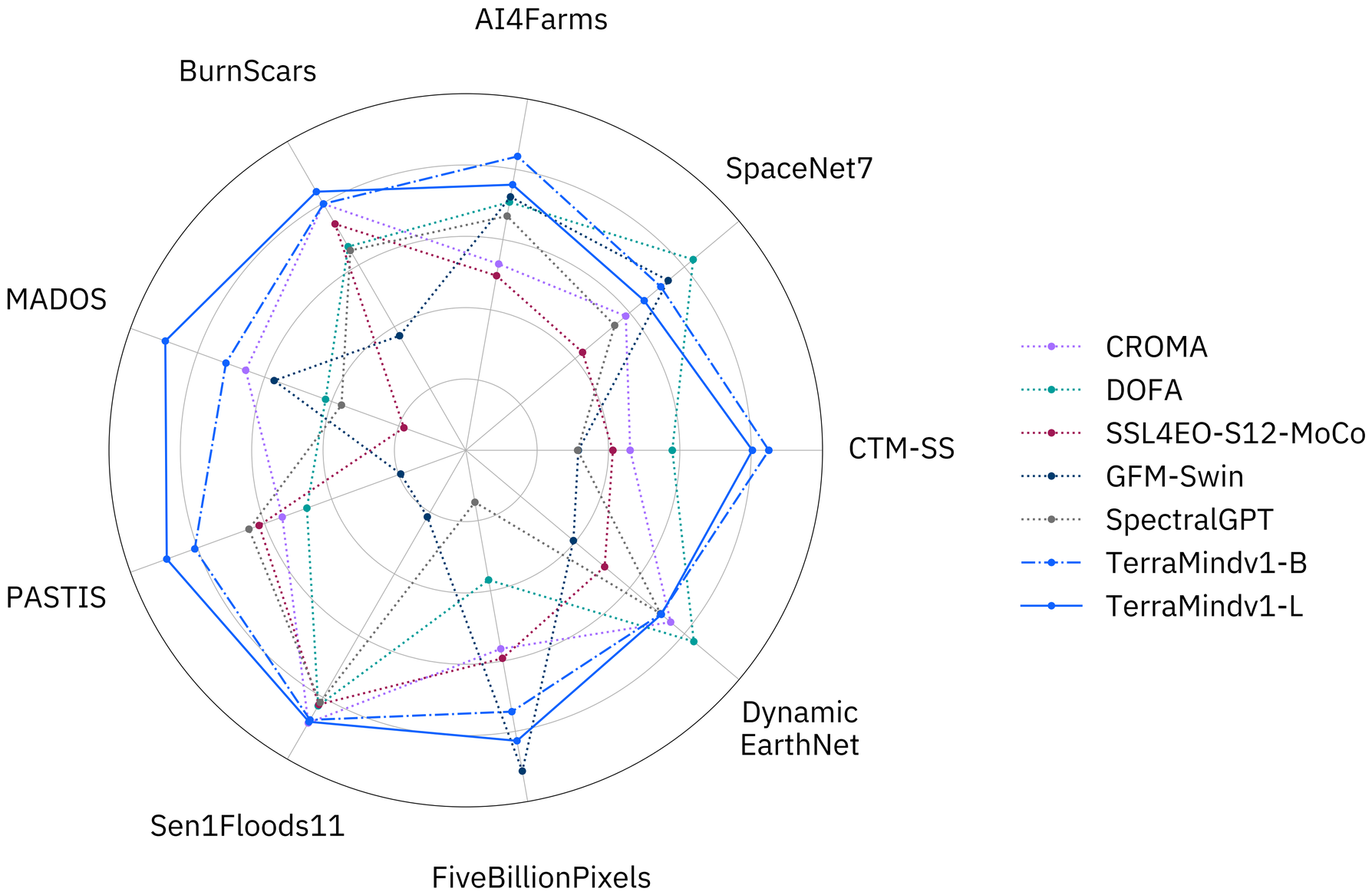
How does TerraMind perform?

PANGAEA bench – nine diverse downstream tasks



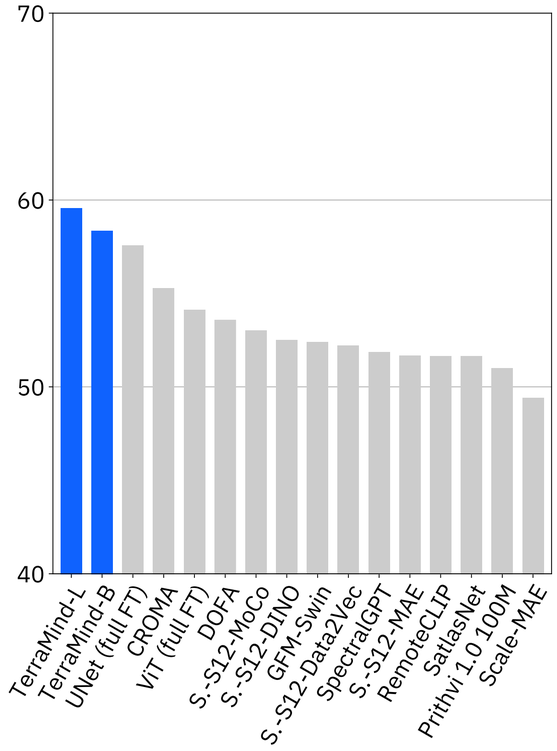
Marsocci, V., Jia, Y., Bellier, G. L., Kerekes, D., Zeng, L., Hafner, S., ... & Nascetti, A. (2024). PANGAEA: A global and inclusive benchmark for geospatial foundation models. *arXiv preprint arXiv:2412.04204*.

PANGAEA bench results



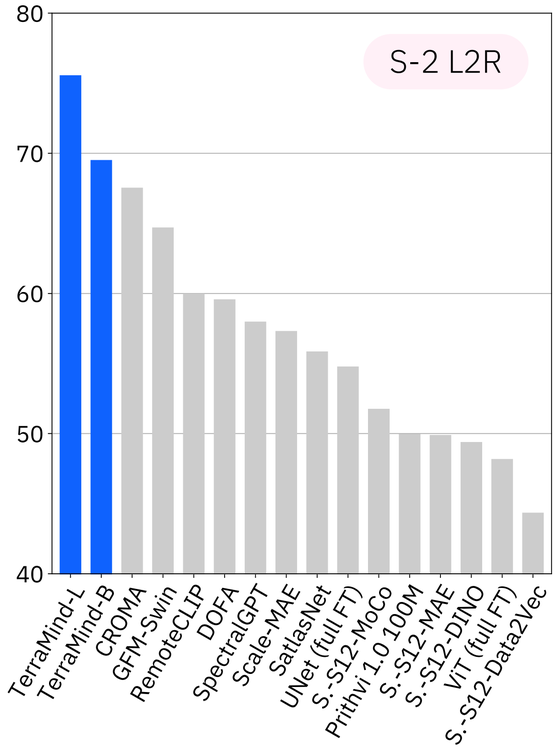
PANGAEA bench results for TerraMind and the top 5 EO FMs based on average rank. The mIoU is visualized on a normalized scale.

Average mIoU



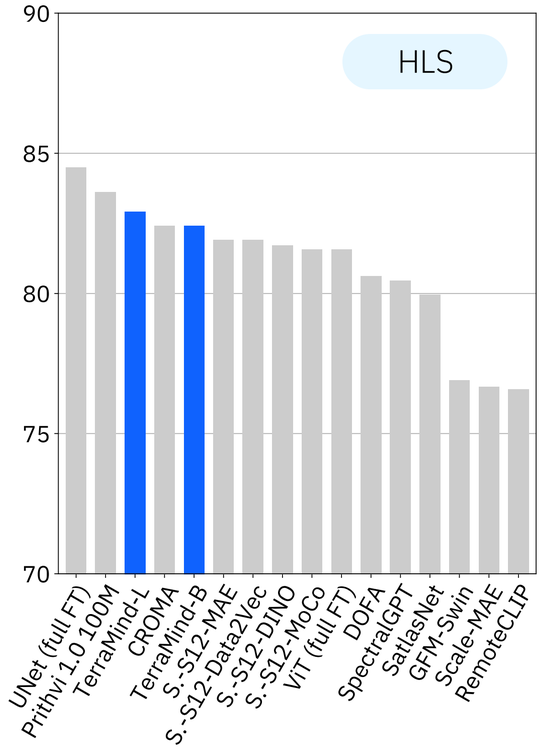
TerraMind is the first EO FM to outperform a fully fine-tuned UNet.

MADOS



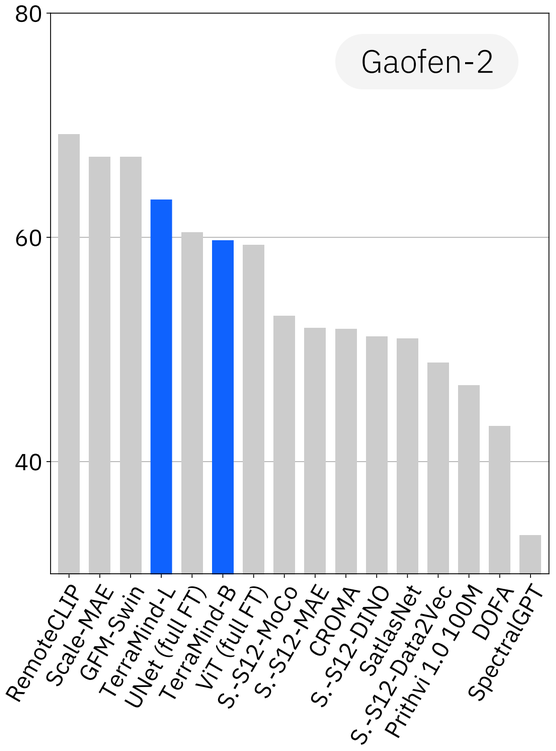
Very good performance on pre-training input modalities like S-2.

Burn Scars



Competitive results on domain shifts like different inputs (HLS).

FiveBillionPixels

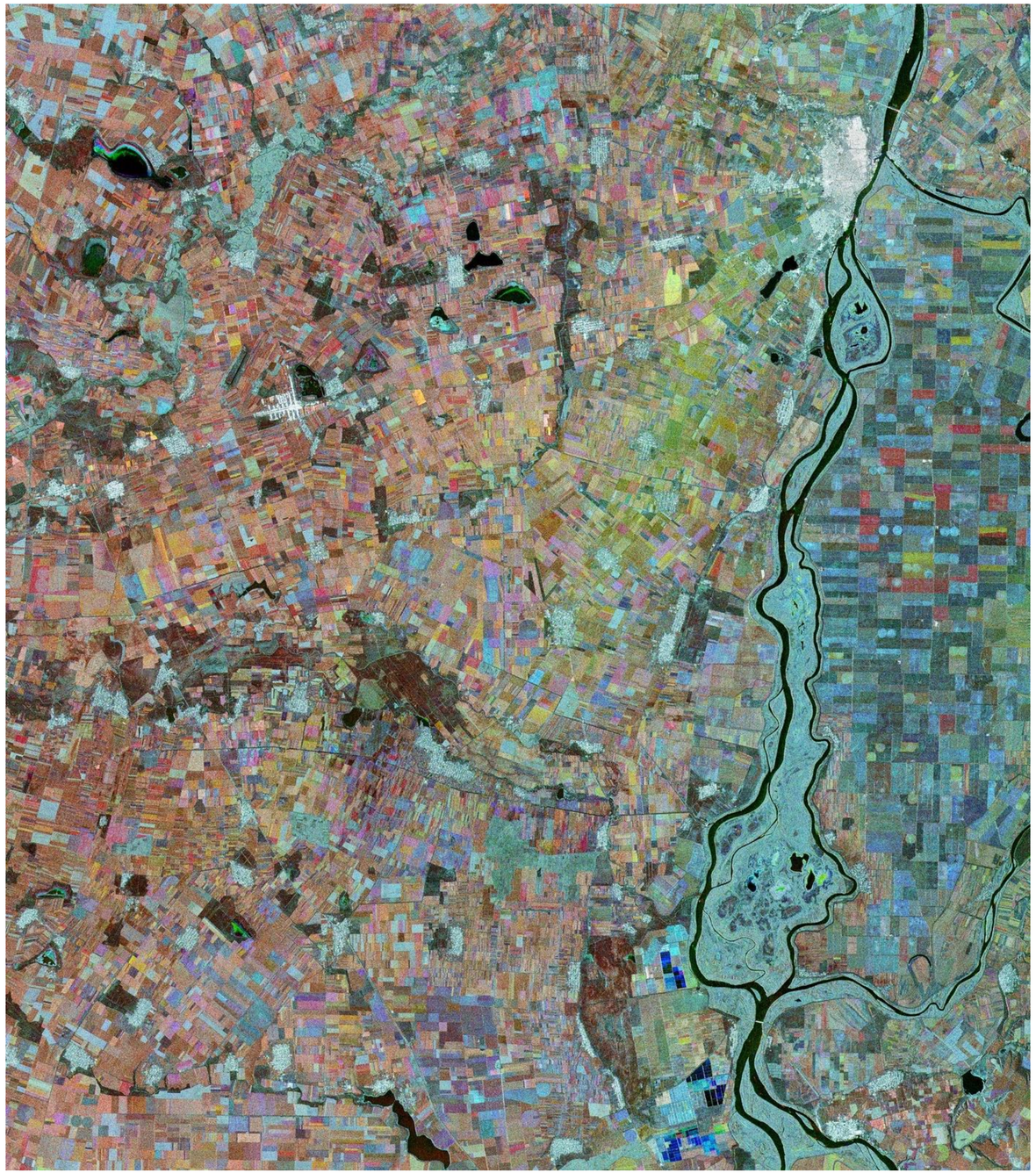


High resolution inputs are challenging for most geospatial FMs.

Improvements with multi-modality

Multimodal inputs consitently improve the TerraMind performance on PANGAEA benchmark datasets.

Input	PASTIS	Sen1Fl11	CTM-SS
Sentinel-1	20.04	80.39	24.45
Sentinel-2	40.20	89.57	50.90
Multimodal	**40.51**	**90.62**	**55.80**

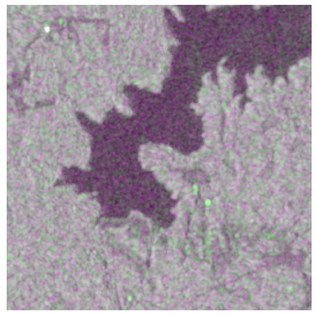


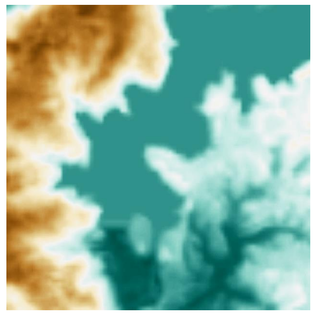
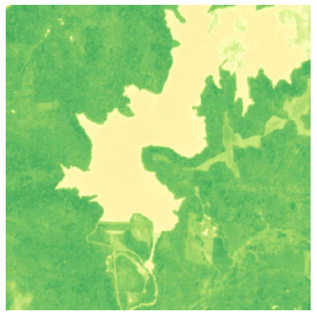
Thinking in Modalities

Comparision between the standard approach with full fine-tuning and Thinking-in-Modalities (TiM) tuning using generated LULC tokens as additional inputs.

Dataset	Model	Input	IoU_{Water}	mIoU
Sen1Floods11	TerraMind-B	Sentinel-1	68.00	81.06
	TerraMind-B TiM	S-1 + *gen. LULC*	**72.25**	**83.65**
	TerraMind-B	Sentinel-2	82.26	89.70
	TerraMind-B TiM	S-2 + *gen. LULC*	**84.75**	**91.14**
SA Crop Type	TerraMind-B	Sentinel-2	–	41.87
	TerraMind-B TiM	S-2 + *gen. LULC*		**42.74**

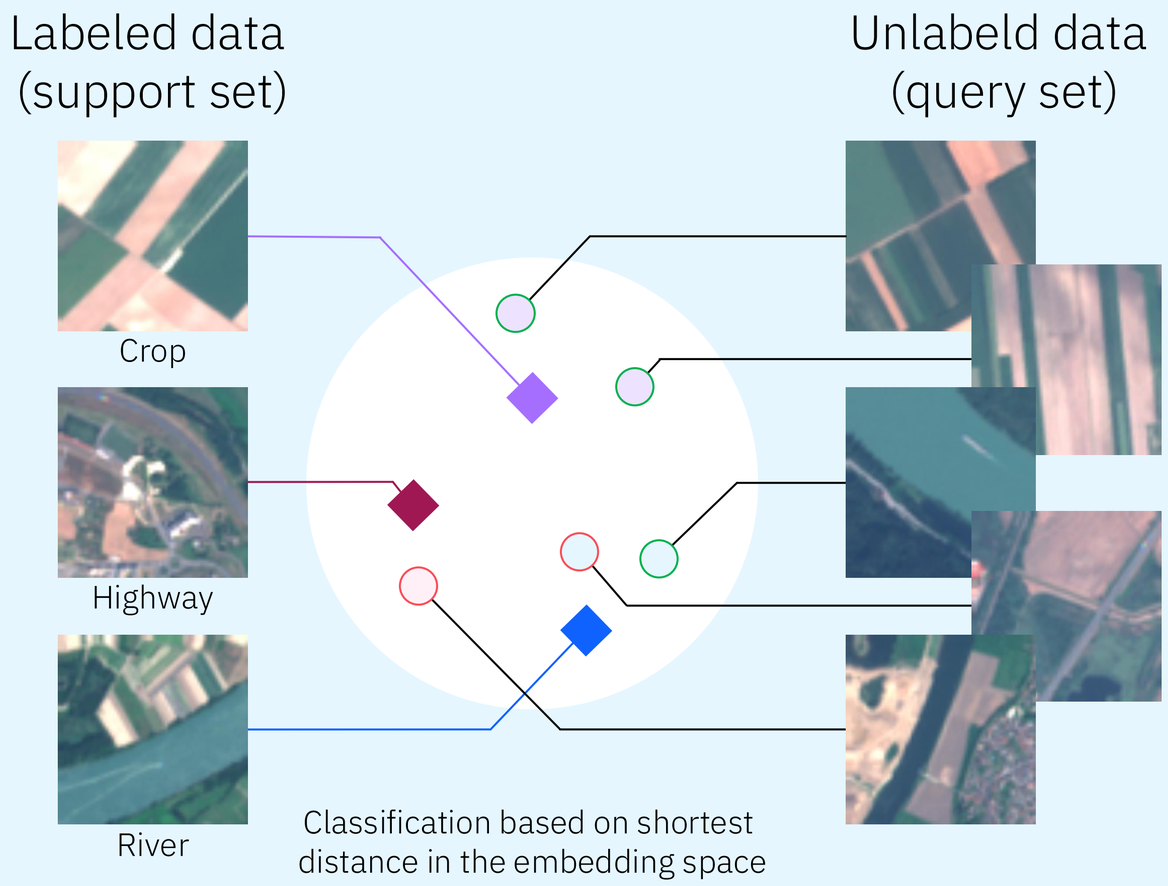
Potential TiM modalities for TerraMind include S-2, S-1, LULC, DEM, NDVI, and coordinates.

Few-shot experiments

How one-shot classification works:



1-shot 5-way classification results using nearest neighbors, measured in accuracy and averaged over 200 runs. TerraMind outperforms benchmarks from CV and EO, suggesting a well-structured latent space.

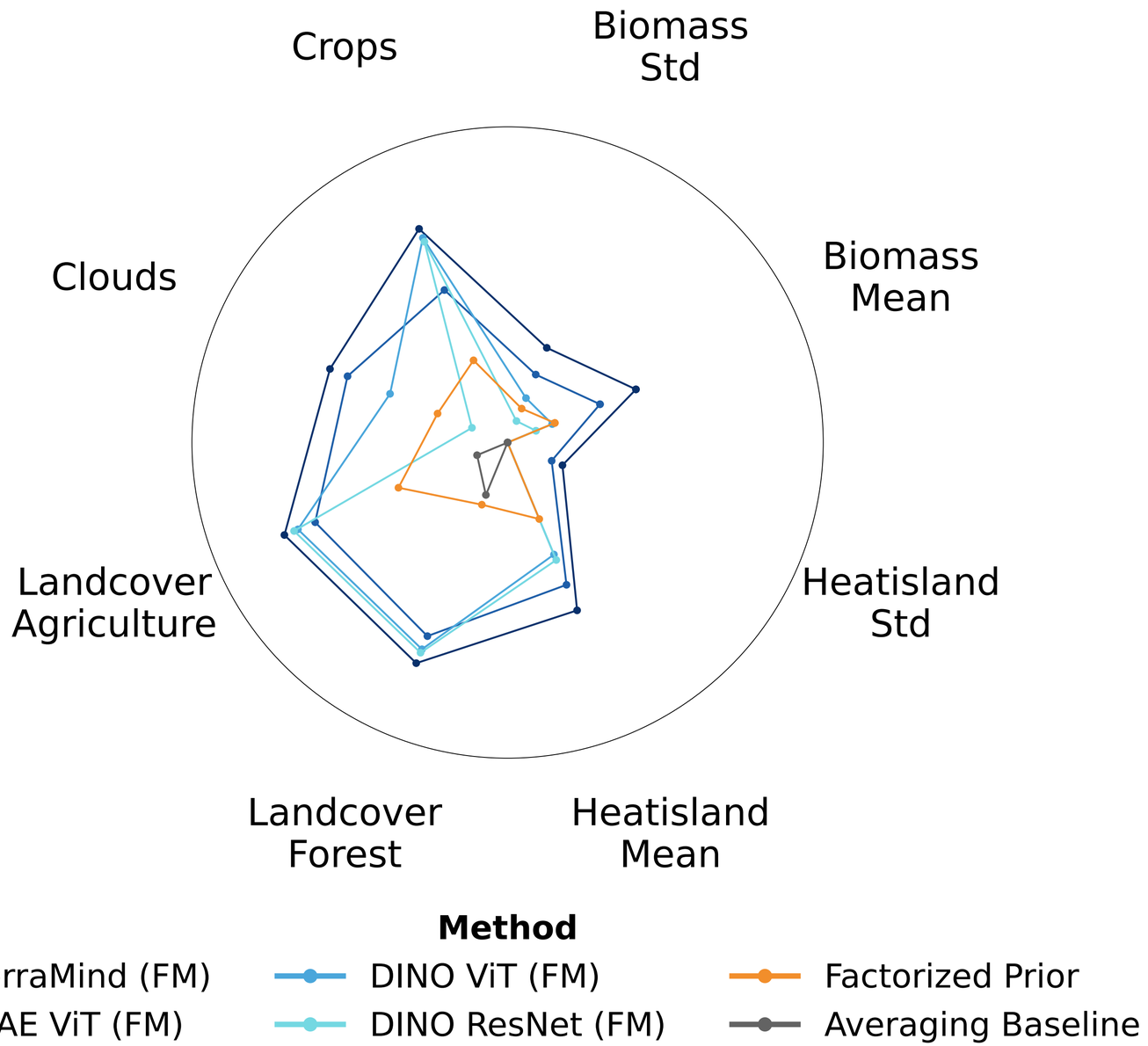
Model	Input	EuroSAT	METER-ML
CLIP-ViT-B/16	S-2RGB	52.39	28.13
	NAIP	–	31.73
DeCUR	S-2 L1C	50.54	27.87
Prithvi EO 1.0	S-2 L1C	60.11	26.08
Prithvi EO 2.0	S-2 L1C	61.06	28.26
TerraMind-B	S-2 L1C	**70.83**	**33.90**
	NAIP	–	32.23

1-shot 5-way defines a setting with five randomly sampled classes with one labeled example each.

Neural compression benchmark

R^2 results of TerraMind, other FMs and baselines on image compression tasks from a CVPR EarthVision challenge.

The well structured embedding space of TerraMind outperforms all other tested models.



Insights

SOTA Performance

TerraMind reaches state-of-the-art performance on PANGAEA bench.

Multi-modal inputs

TerraMind is trained with various modalties, natively supporting multi-modal fine-tuning.

Thinking-in-Modalites

TerraMind introdcues a new intermediate step and imagines how other modalities look like.

Embedding space

TerraMind embeddings are well well suited for few-shot classification or image retrival tasks.

Maritime use case with TerraMind

Welcome Benedikt Blumenstiel!

Manage you API keys →

Manage your API keys to access the studio services

Documentation →

Access SDKs, model cards, and developer resources

Feedback →

Manage your feedback submissions and check for status

Prepare data

Refine datasets for tuning and labeling

Open Data Factory →

Customize a model

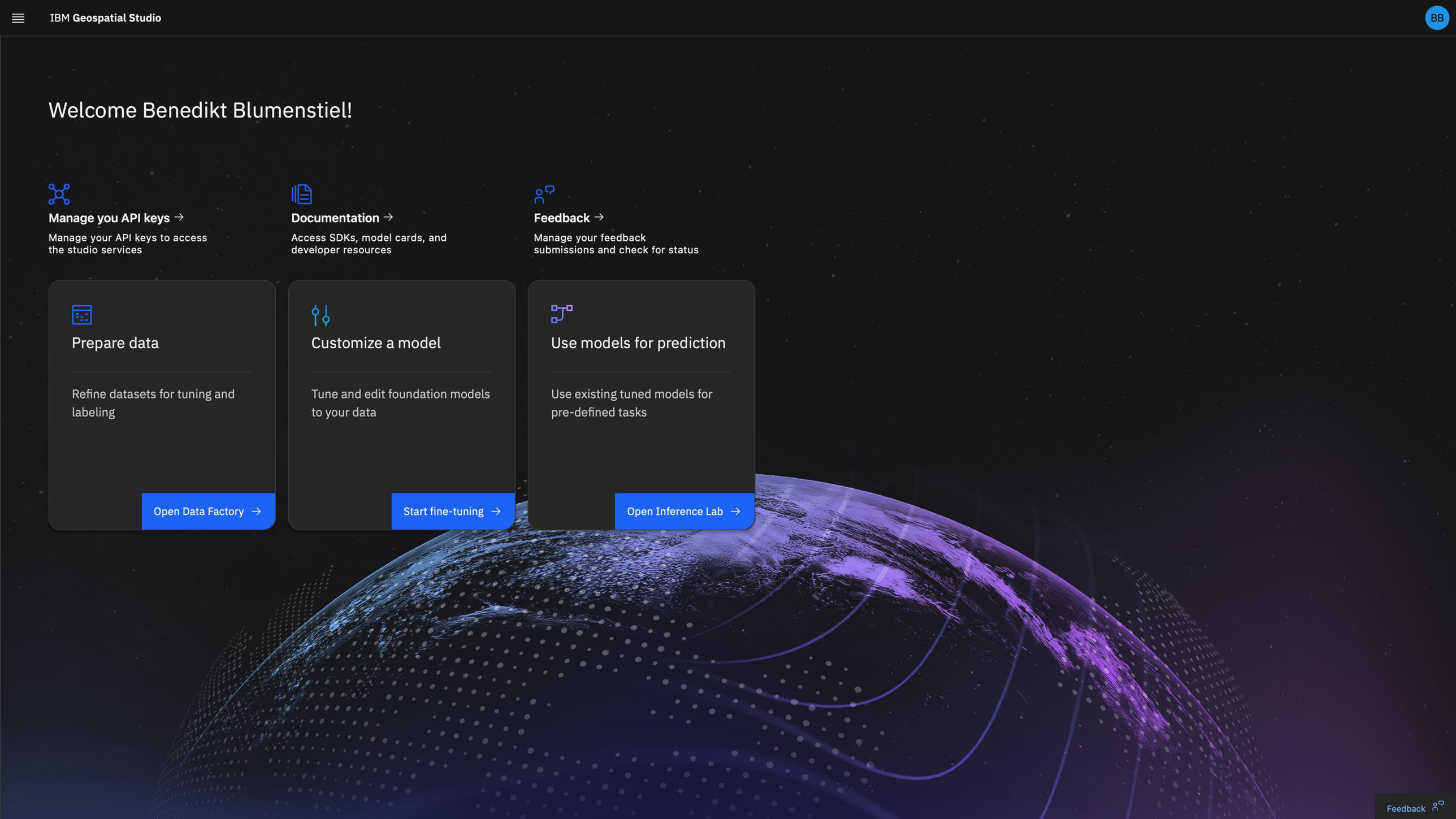
Tune and edit foundation models to your data

Start fine-tuning →

Use models for prediction

Use existing tuned models for pre-defined tasks

Open Inference Lab →



New_York - Activity-led maritime monitoring

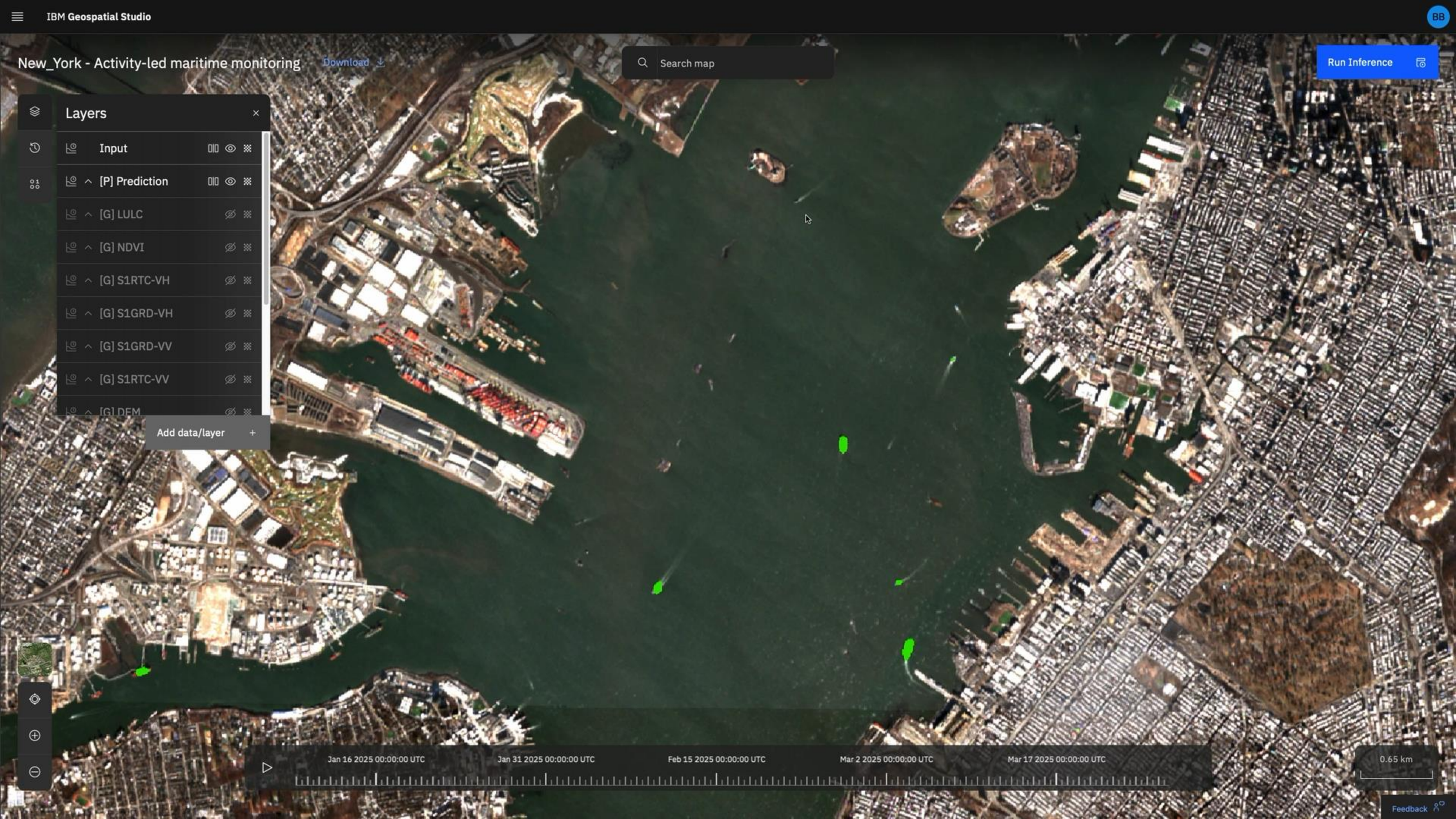
Download

Search map

Run Inference

Jan 16 2025 00:00:00 UTC | Jan 31 2025 00:00:00 UTC | Feb 15 2025 00:00:00 UTC | Mar 2 2025 00:00:00 UTC | Mar 17 2025 00:00:00 UTC

Feedback


IBM Geospatial Studio
BB
New_York - Activity-led maritime monitoring
Download
Search map
Run Inference
Layers
Input
[P] Prediction
[G] LULC
[G] NDVI
[G] S1RTC-VH
[G] S1GRD-VH
[G] S1GRD-VV
[G] S1RTC-VV
[G] DEM
Add data/layer
Jan 16 2025 00:00:00 UTC
Jan 31 2025 00:00:00 UTC
Feb 15 2025 00:00:00 UTC
Mar 2 2025 00:00:00 UTC
Mar 17 2025 00:00:00 UTC
0.65 km
Feedback


IBM Geospatial Studio
BB
New_York - Activity-led maritime monitoring
Download
Search map
Run Inference
Layers
[G] S1RTC-VH
[G] S1GRD-VH
[G] S1GRD-VV
[G] S1RTC-VV
[G] DEM
[P] Prediction TiM
[P] Monitoring Map
Harbour Activity
3D buildings
Add data/layer
Boston
Houston
Los_Angeles
New_York
San_Francisco
Seattle
Q1 2025 Activities
1.5
1.0
0.5
0.0
:00 UTC
Jan 16 2025 00:00:00 UTC
Jan 31 2025 00:00:00 UTC
Feb 15 2025 00:00:00 UTC
Mar 2 2025 00:00:00 UTC
Mar 17 2025 00:00:00 UTC
Apr 1 202
3.34 km
Feedback

Leveraging TerraMind in Singapore

Singapore - Activity-led Maritime Monitoring

Download

Search map

Run Inference

Jan 16 2025 00:00:00 UTC | Jan 31 2025 00:00:00 UTC | Feb 15 2025 00:00:00 UTC | Mar 2 2025 00:00:00 UTC | Mar 17 2025 00:00:00 UTC | Apr 1

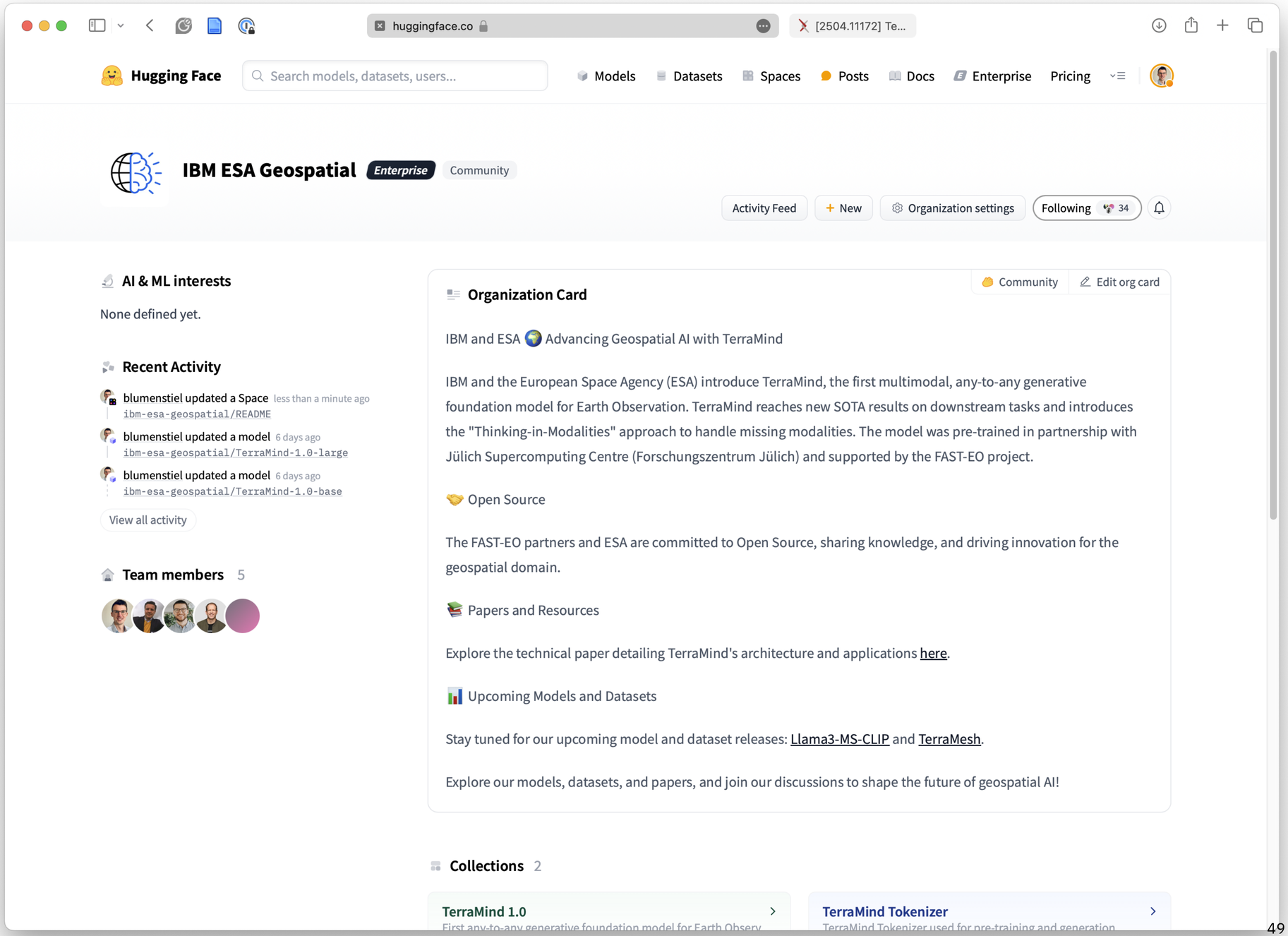
Feedback

Try it for yourself!

We ♥ Open source!

All models are released on Hugging Face under the Apache 2.0 license.

https://huggingface.co/ibm-esa-geospatial



TerraTorch – The FM Toolkit

TerraMind is fully integrated into TerraTorch which enables low/no-code fine-tuning.

Initalize pre-trained TerraMind models for fine-tuning, TiM tuning, any-to-any generation, or tokenization within a few lines of code.

```
from terratorch import BACKBONE_REGISTRY

model = BACKBONE_REGISTRY.build(
    "terramind_v1_base",
    modalities=["S2L1C", "S1GRD"],
    pretrained=True,
)
```

Examples at https://github.com/IBM/terramind

We want to see what you can build with TerraMind!

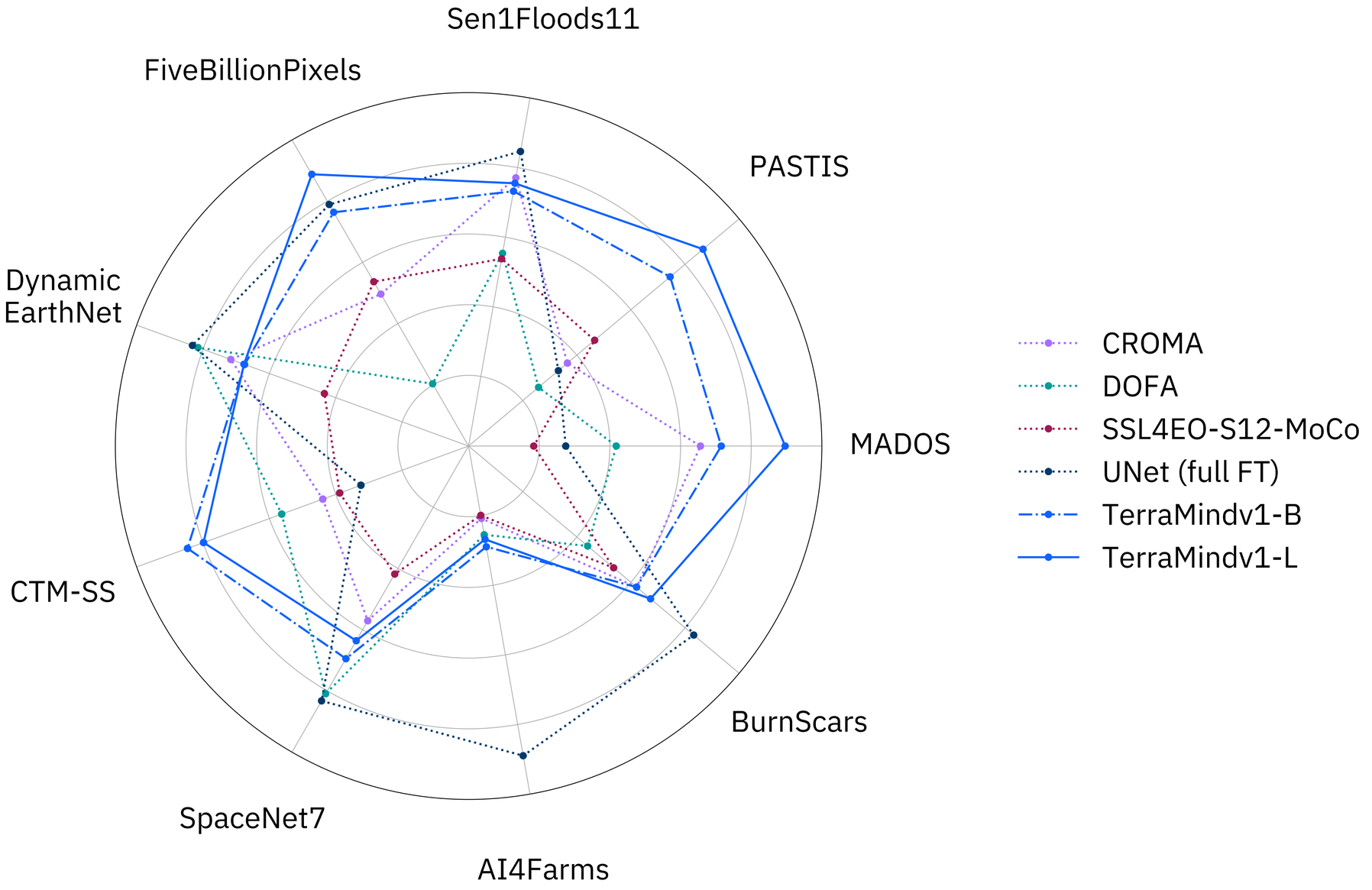
Get inspired and submit your idea with TerraMind at:
https://huggingface.co/ibm-esa-geospatial/challenge

A bi-monthly award spotlighting the boldest, most imaginative ways to push TerraMind beyond “just another fine-tune”. You can win 4 x €1 000 prize money.

See Terms & Conditions at https://huggingface.co/ibm-esa-geospatial/challenge.

Thank you!

Find more information about TerraMind at
https://huggingface.co/ibm-esa-geospatial



PANGAEA bench results for TerraMind, the top 3 EO FMs and a fully fine-tuned UNet. The mIoU is visualized on a normalized scale.

PANGAEA bench

Performance evaluation of TerraMind across nine benchmark datasets using the PANGAEA evaluation protocol. Higher mIoU (↑) and lower rank values (↓) are reported. The best model is highlighted and the second best is underscored.

Model	BurnSr	MADOS	PASTIS	Sen1Fl11	FBP	DEN	CTM-SS	SN7	AI4Farms	Avg. mIoU	Avg. Rank
CROMA	82.42	67.55	32.32	90.89	51.83	38.29	49.38	59.28	25.65	55.29	6.61
DOFA	80.63	59.58	30.02	89.37	43.18	39.29	51.33	61.84	27.07	53.59	8.22
GFM-Swin	76.90	64.71	21.24	72.60	67.18	34.09	46.98	60.89	27.19	52.42	10.00
Prithvi 1.0 100M	83.62	49.98	33.93	90.37	46.81	27.86	43.07	56.54	26.86	51.00	11.00
RemoteCLIP	76.59	60.00	18.23	74.26	**69.19**	31.78	52.05	57.76	25.12	51.66	11.22
SatlasNet	79.96	55.86	17.51	90.30	50.97	36.31	46.97	61.88	25.13	51.65	10.67
Scale-MAE	76.68	57.32	24.55	74.13	67.19	35.11	25.42	**62.96**	21.47	49.43	11.44
SpectralGPT	80.47	57.99	35.44	89.07	33.42	37.85	46.95	58.86	26.75	51.87	10.11
S.-S12-MoCo	81.58	51.76	34.49	89.26	53.02	35.44	48.58	57.64	25.38	53.02	10.06
S.-S12-DINO	81.72	49.37	36.18	88.61	51.15	34.81	48.66	56.47	25.62	52.51	10.89
S.-S12-MAE	81.91	49.90	32.03	87.79	51.92	34.08	45.80	57.13	24.69	51.69	12.39
S.-S12-Data2Vec	81.91	44.36	34.32	88.15	48.82	35.90	54.03	58.23	24.23	52.22	10.72
UNet Baseline	**84.51**	54.79	31.60	**91.42**	60.47	**39.46**	47.57	62.09	**46.34**	57.58	4.89
ViT Baseline	81.58	48.19	38.53	87.66	59.32	36.83	44.08	52.57	38.37	54.13	10.28
TerraMindv1-B	82.42	69.52	40.51	90.62	59.72	37.87	**55.80**	60.61	28.12	58.35	3.94
TerraMindv1-L	82.93	**75.57**	**43.13**	90.78	63.38	37.89	55.04	59.98	27.47	**59.57**	**3.44**

Jakubik, J., Yang, F., Blumenstiel, B., Scheurer, E., Sedona, R., Maurogiovanni, S., ... & Longépé, N. (2025). TerraMind: Large-Scale Generative Multimodality for Earth Observation. *arXiv preprint arXiv:2504.11171*.