#EUSpace

E2S EMBED 2SCALE

Earth Observation &Weather Data Federation With AI Embeddings

EUSPA AI week 2026

Thomas Brunschwiler

IBM Research Europe

EUSPA European Union Agency for the Space Programme

Data Gravity in Earth & Weather Data

10 years

Global flood mapping with Sentinel-1&2.

22PB

Data volume per modality.

eGres $8Mio

… and 520 days download time at 10GBit/s.

Compute $280k

… for TerraMind on A100 requires 3'000 GPU/days.

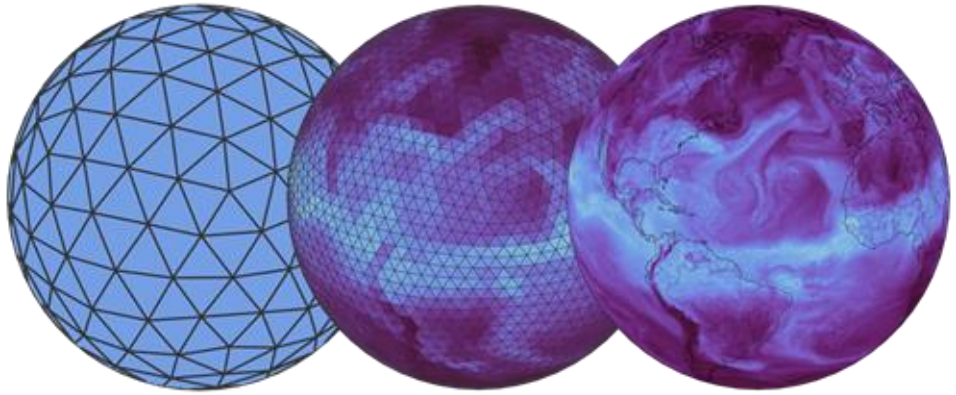


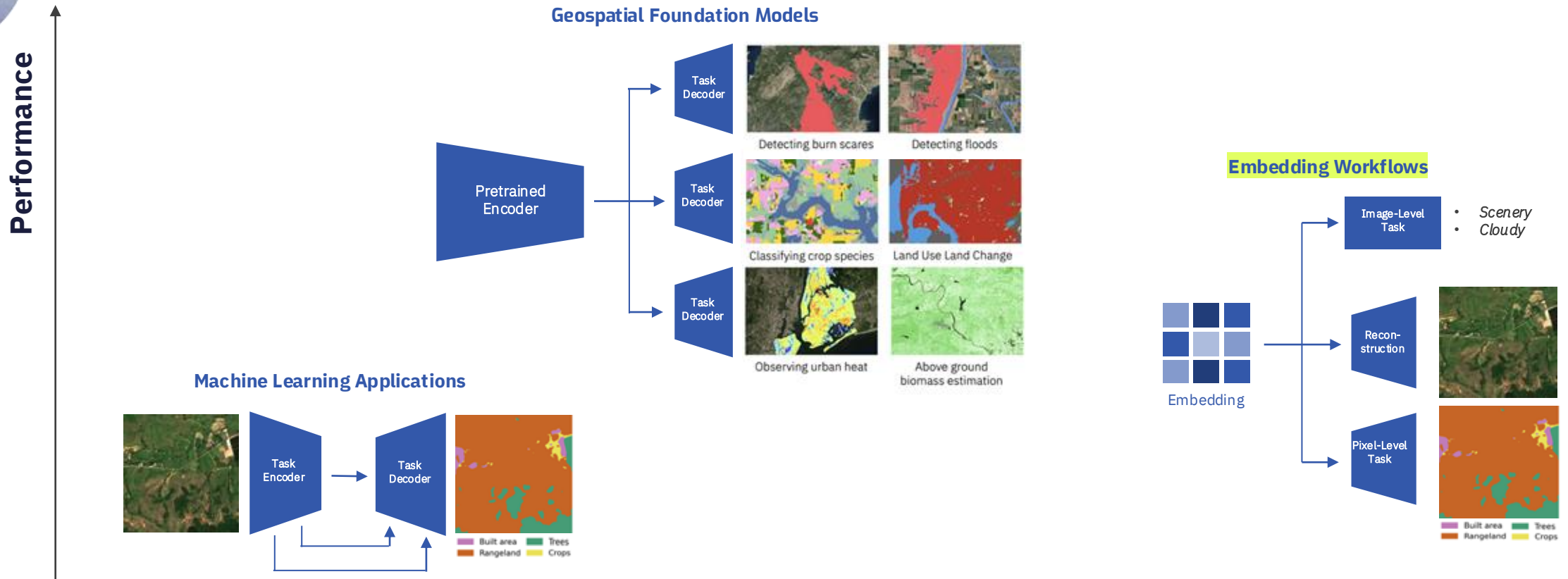
Image source: https://easy.gems.dkrz.de/

ECMWF

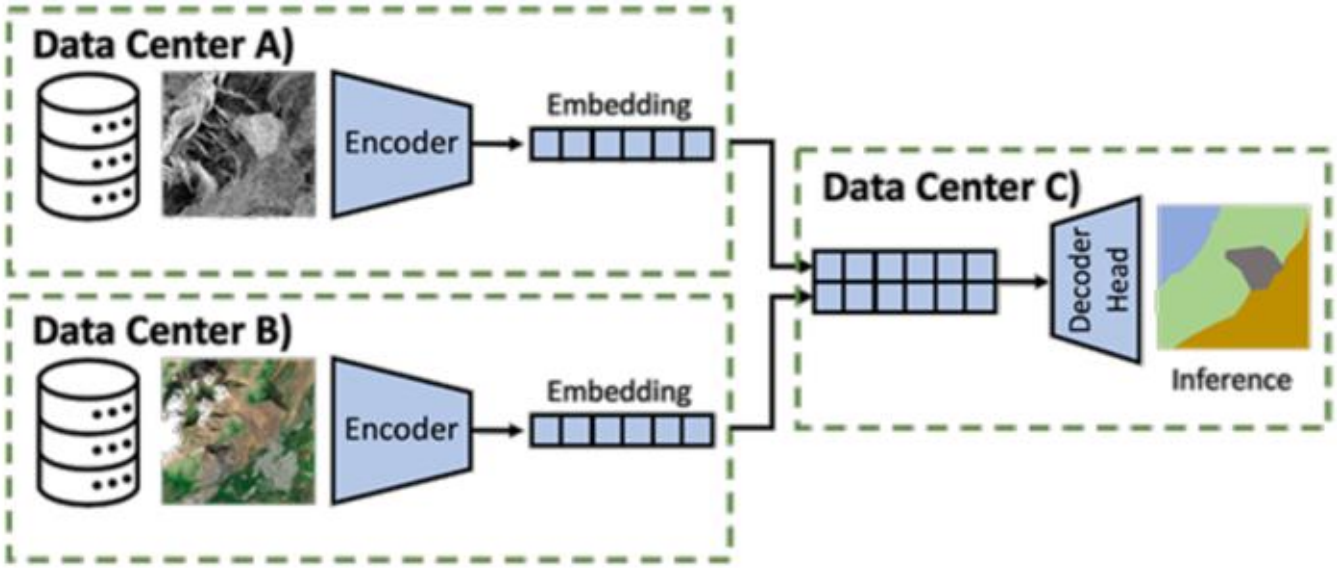
Data archive reaches one exabyte

ECMWF Annual Report 2024

Evolution of AI in Earth Observation



Embedding Federation



Embed2Store

Computer once, request many times

Embed2Infer

Light-weight decoders and reduced compute

Embed2Transfer

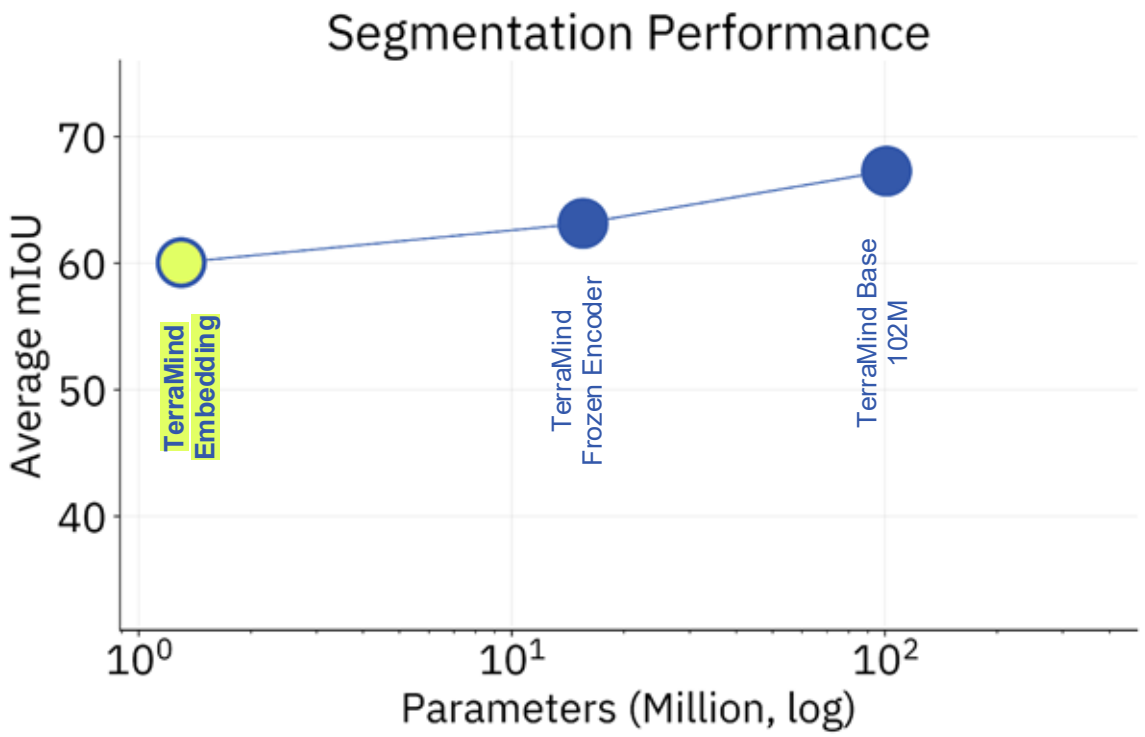
Decentralised applications

Embed2Find

Image retrieval from petabyte scale data

Meet the E2S EMBED 2SCALE Consortium

Performance of Embedding Workflows



Embeddings need

- **200x fewer** modelling parameters
- **4x smaller** data size

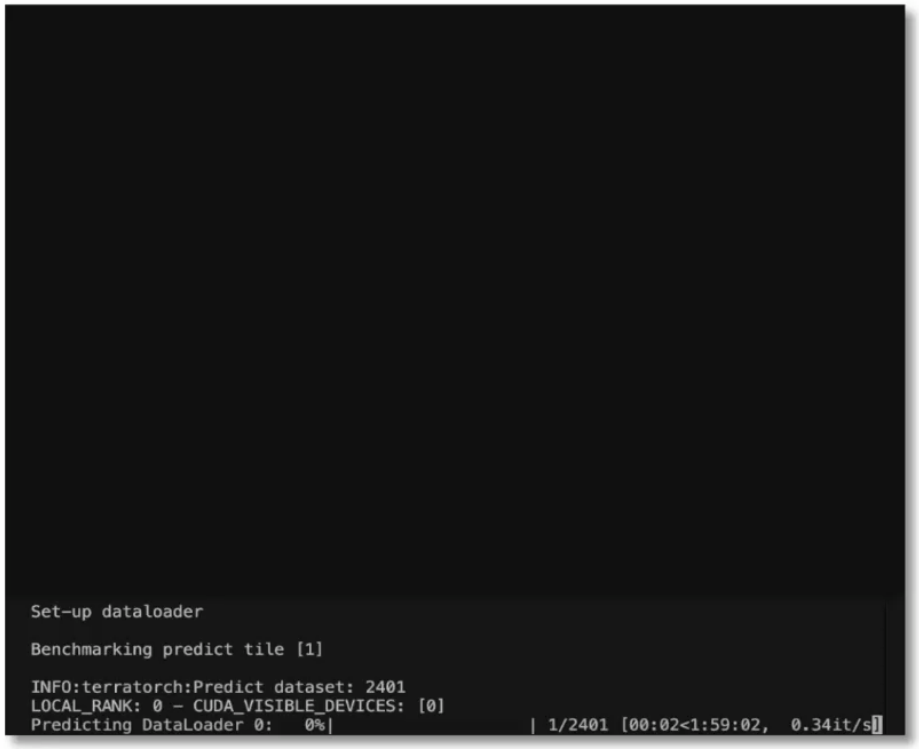
Average over 3 Segmentation Task: HLS Burnscars, Sentinel-2 Scene Classification Layer, FLAIR2

Inference Speed-Up of Embedding Workflows

Embedding Workflow

1 Tile Inference in 22s

Foundation Model Workflow



1 Tile Inference in 6min

On a A100 GPU, embeddings and raw images are both read in as single file tiffs

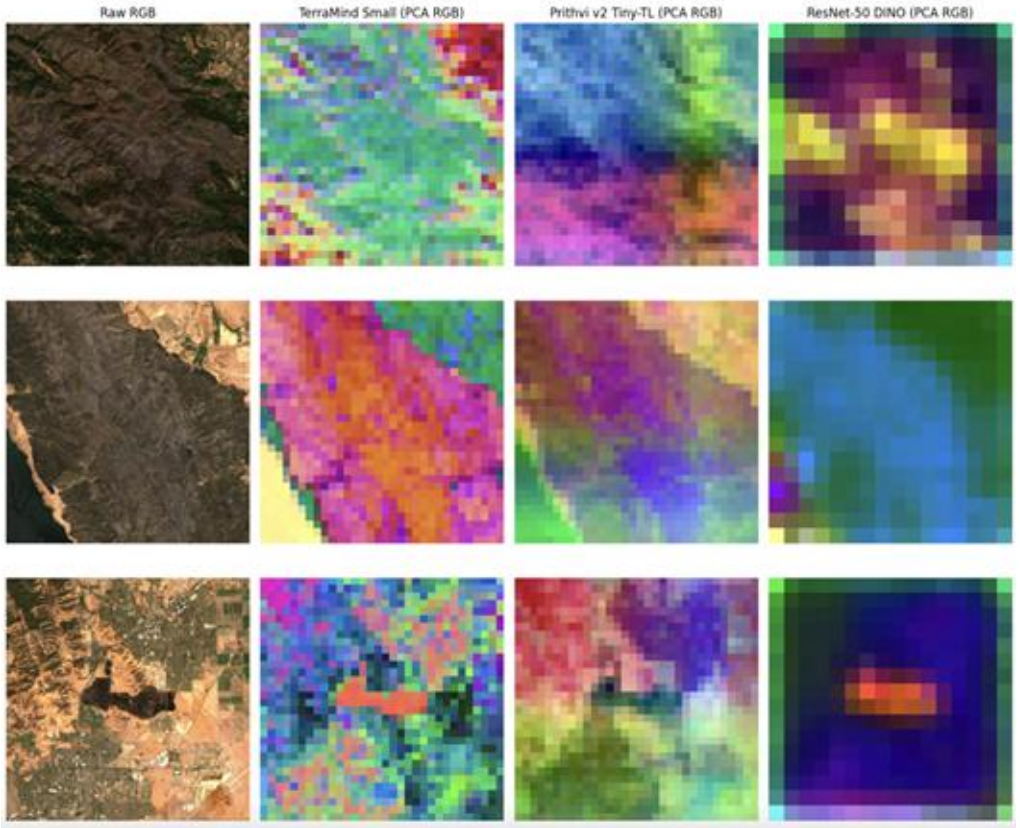
TerraTorch Extension for Embedding Workflows

No-/Low-Code Embedding Generation

```
model = terratorch.tasks.EmbeddingGenerationTask(
    model='prithvi_eo_v2_300',
    model_args={...},

    output_format='tiff',
    output_dir=dataset_path / 'embeddings/',

    layers=[-1],
    embedding_pooling=None,
    has_cls=True
)
trainer = pl.Trainer(
    accelerator="auto",
    strategy="auto",
    max_epochs=0
)

# Generate Embeddings
trainer.predict(model, datamodule=datamodule)
```

[26]

PCA Embeddings of Different Backbones



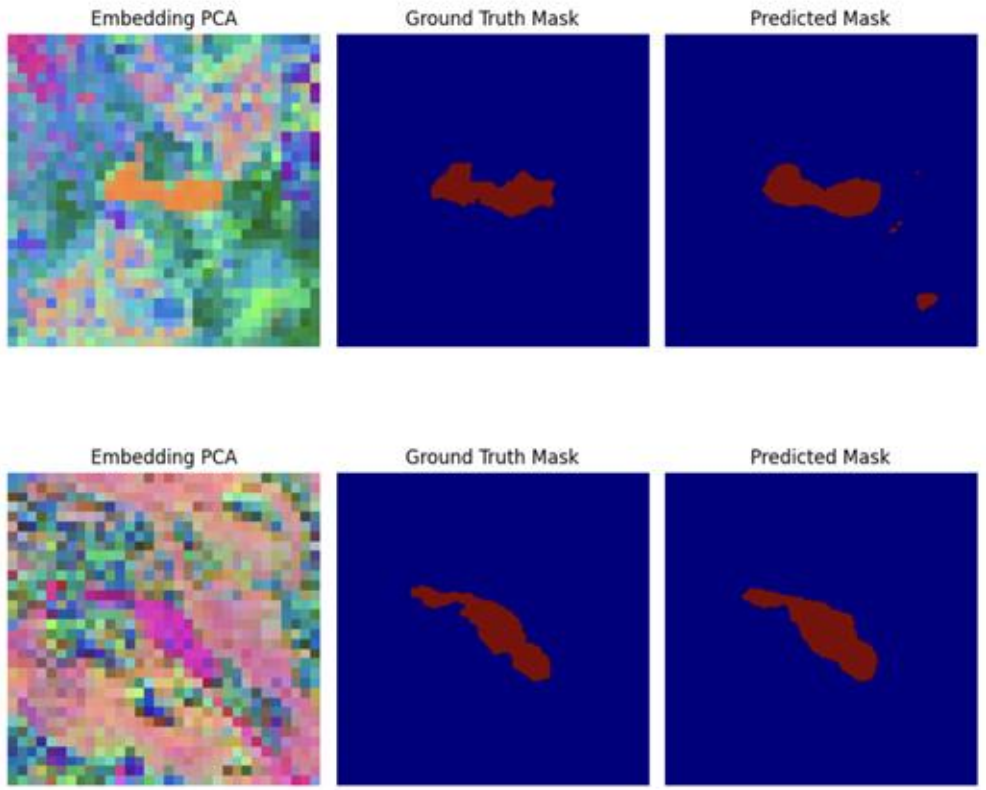
https://github.com/terrastackai/terratorch/tree/main/examples

TerraTorch Extension for Embedding Workflows

Embedding Downstream Tasks

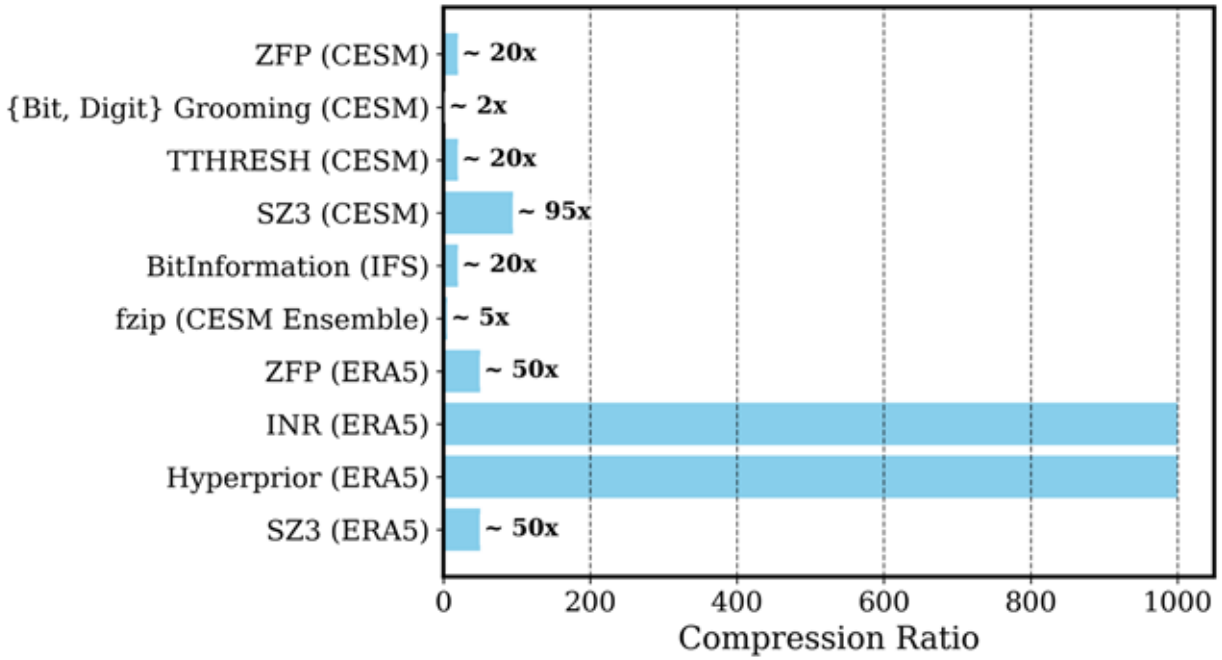
```
model = terratorch.tasks.SemanticSegmentationTask(
    model_factory="EncoderDecoderFactory",
    model_args={
       "backbone": "IdentityBackbone",
        "backbone_out_channels": [768],
        "necks": [
            {...}
        ],
        "decoder": "FCNDecoder",
        "decoder_channels": 128,
        "head_dropout": 0.1,
        "head_channel_list": [128],
        "num_classes": 2,
    },

    loss="ce",
    optimizer="AdamW",
    lr=1e-4,
    freeze_decoder=False,
    class_names=['no burned', 'burned'])
```

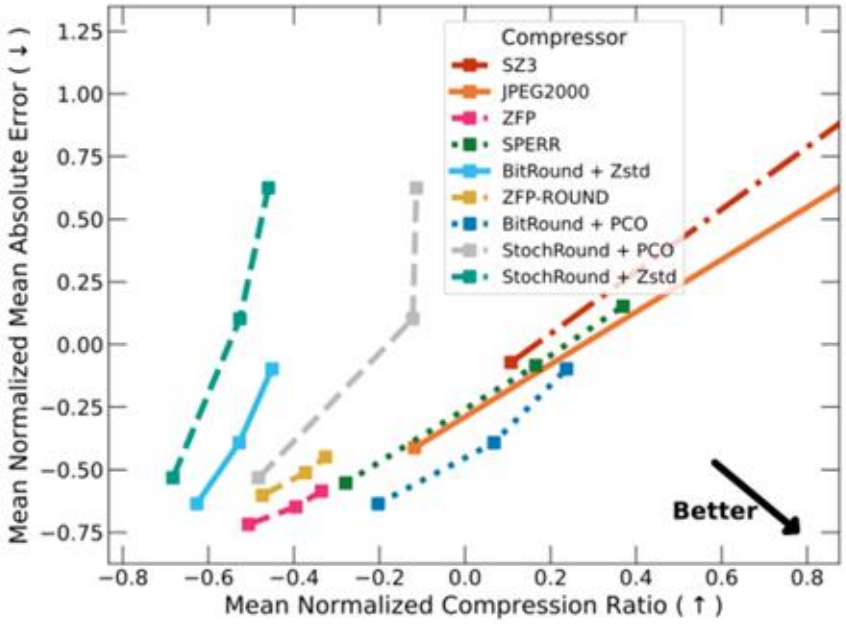
Prediction Performance after Two Epoch Training



https://github.com/terrastackai/terratorch/tree/main/examples

ClimateBenchPress for Weather and Climate Models





Compression ratios, variables, error thresholds and data **varies** in published literature

https://github.com/climatebenchpress

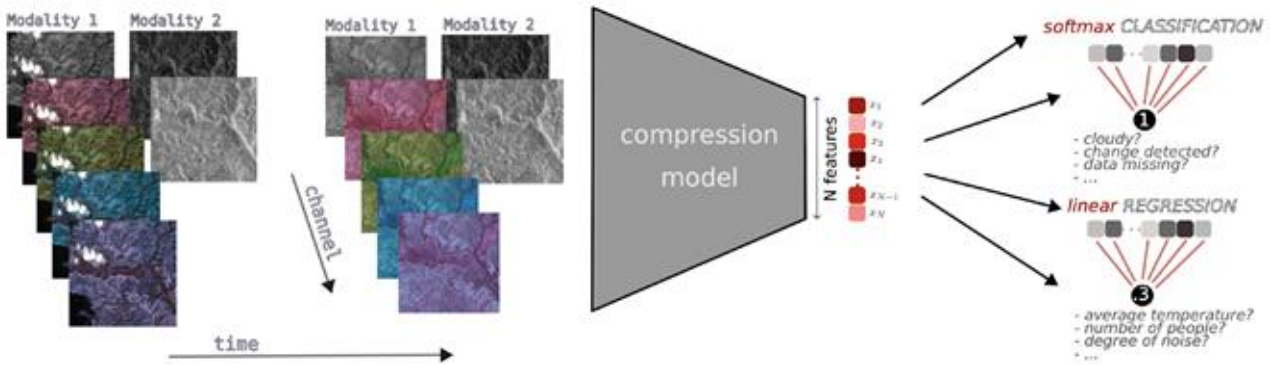
NeuCo-Bench: Benchmarking Earth Observation Embeddings

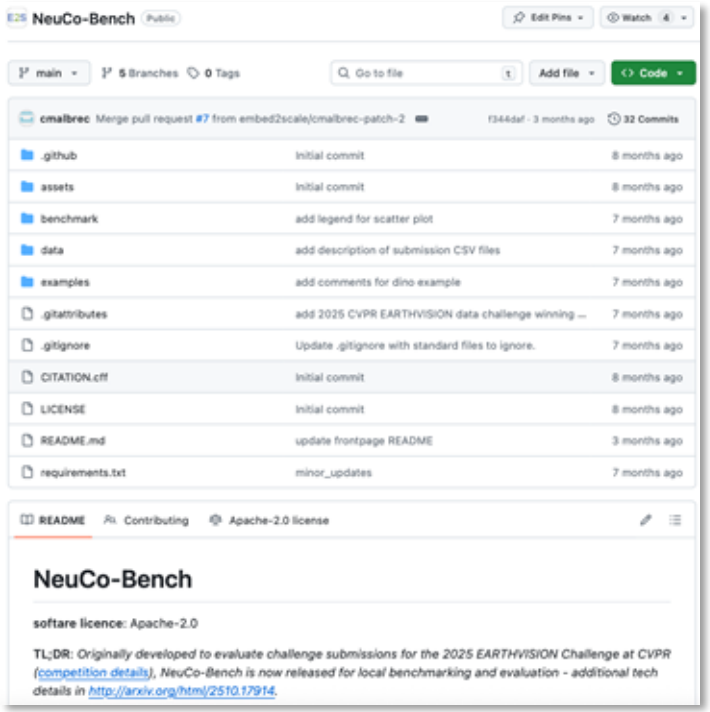
EARTH VISION

Participant Teams
23

Total Submissions
465

~7000x compression





https://github.com/embed2scale/NeuCo-Bench

Earth2Vec Community

CLARK UNIVERSITY
CENTER FOR GEOSPATIAL ANALYTICS

Universität Münster

DLR Deutsches Zentrum für Luft- und Raumfahrt
German Aerospace Center

JÜLICH Forschungszentrum
JÜLICH SUPERCOMPUTING CENTRE

CloudFerro

Spatial Informatics Group

ETH zürich

developmentSEED

Earth Genome

planet.

IBM Research Europe

COLUMBIA UNIVERSITY

eurac research

LEGEND

Degas

Element 84

OroraTech

Universität Zürich[UZH]

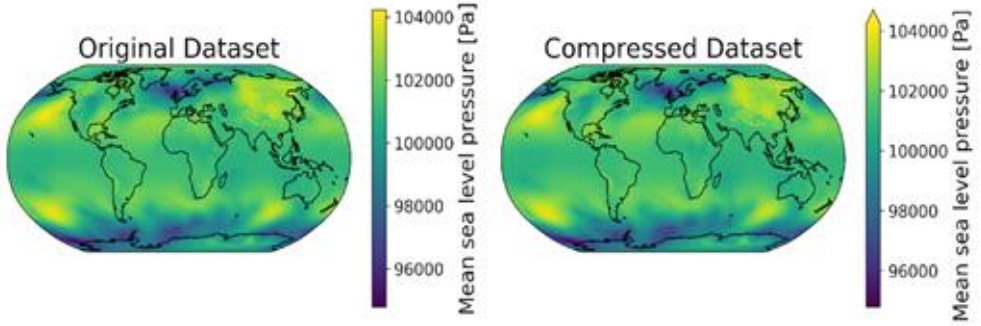
asterisk labs

der Bundeswehr Universität München

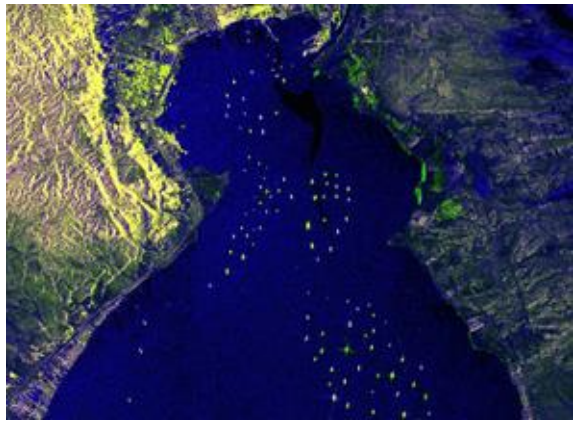
Earth2Vec

Outlook Embedding Use Cases

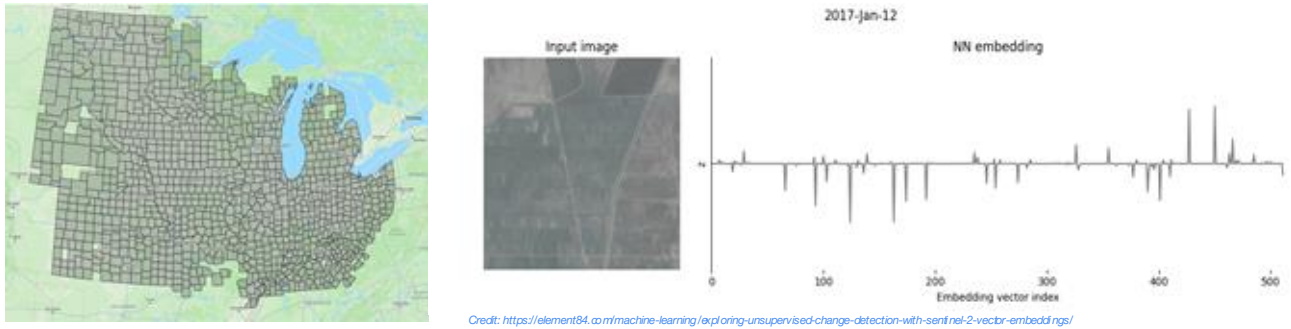
Above Ground Biomass
Embedding Upsampling for Dense Predictions



Climate and Air Pollution Prediction
Next-Gen Climate Model Evaluations in Embedding Spaces



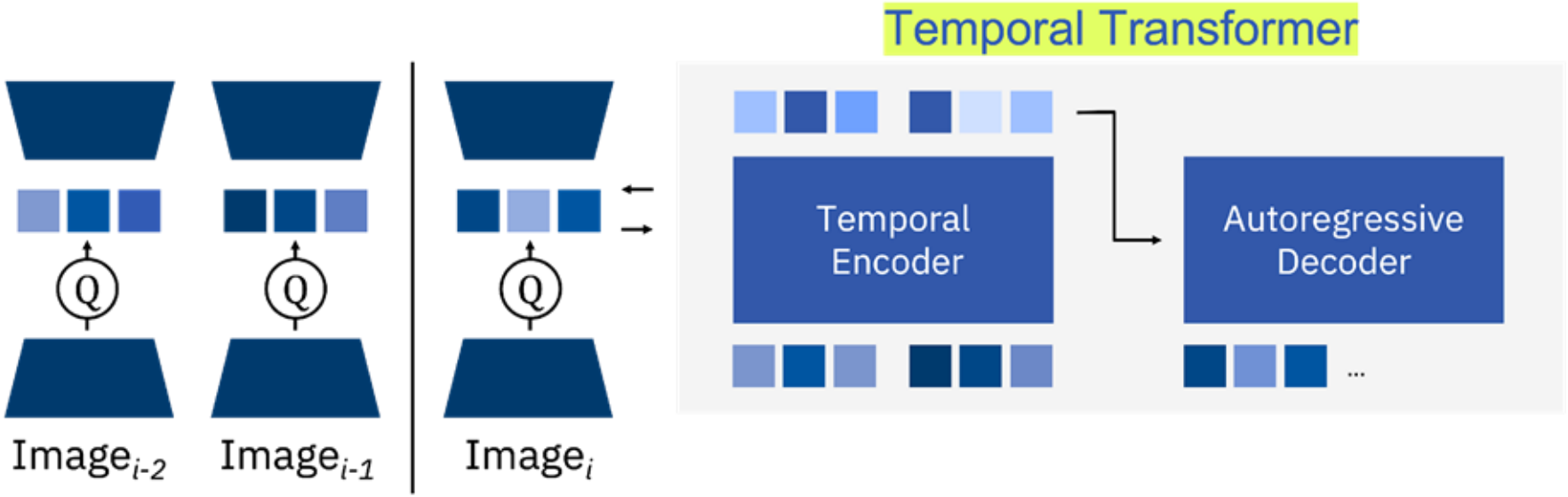
Maritime Awareness
Compressed SAR Data Vessel Detection



Credit: https://element84.com/machine-learning/exploring-unsupervised-change-detection-with-sentinel-2-vector-embeddings/

Crop Stress and Yield Prediction
TerraMind Embeddings for CDSE

Outlook Learned Temporal Compression for EO



TerraCodec Release:

First public neural codecs pretrained **on multi-spectral & multi-temporal** Sentinel-2 data.

#EUSpace

EUSPA
European Union Agency for the Space Programme

Linking space to user needs

Get in touch with us

www.euspa.europa.eu

EUSPA EU4Space EUSPA space4eu EUSPA

EUSPA AI WEEK 2026

E2S EMBED 2SCALE

Lossy Neural Compression for Geospatial Analytics

A review

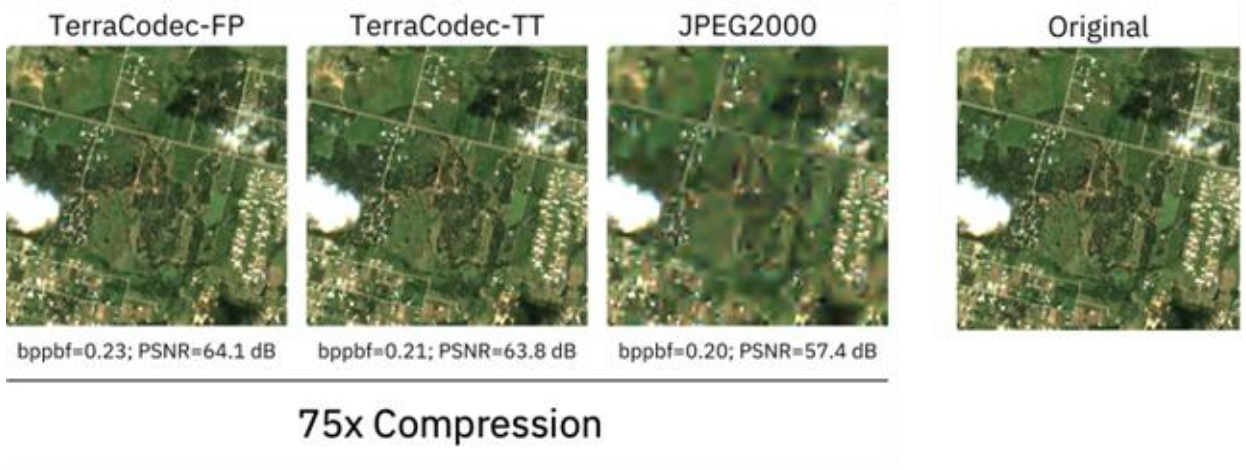
CARLOS GOMES, ISABELLE WITTMANN, DAMIEN ROBERT, JOHANNES JAKUBIK, TIM REICHELT, STEFANO MAUROGIOVANNI, RIKARD VINGE, JONAS HURST, ERIK SCHEURER, ROCCO SEDONA, THOMAS BRUNSCHWILER, STEFAN KESSELHEIM, MATEJ BATIČ, PHILIP STIER, JAN DIRK WEGNER, GABRIELE CAVALLARO, EDZER PEBESMA, MICHAEL MARSZALEK, MIGUEL A. BELENGUER-PLOMER, KENNEDY ADRIKO, PAOLO FRACCARO, ROMEO KIENZLER, RANIA BRIQ, SABRINA BENASSOU, MICHELE LAZZARINI, AND CONRAD M. ALBRECHT

Published in *IEEE Geoscience and Remote Sensing Magazin, 2025*

Outlook TerraCodec Results

3-10x lower bitrate at equal image quality

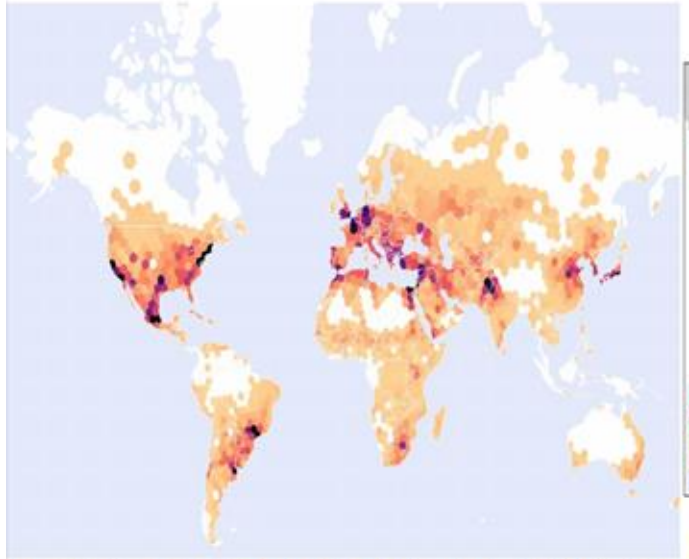
Up to 380× compression with <1pp drop in task performance



	Sen1Floods11	
Compression	**mIoU**	**IoU$_{Flood}$**
Original data	87.77	78.75
170×	87.31 (-0.46)	78.02 (-0.73)
380×	87.27 (-0.50)	77.97 (-0.78)
940×	86.76 (-1.01)	77.06 (-1.69)

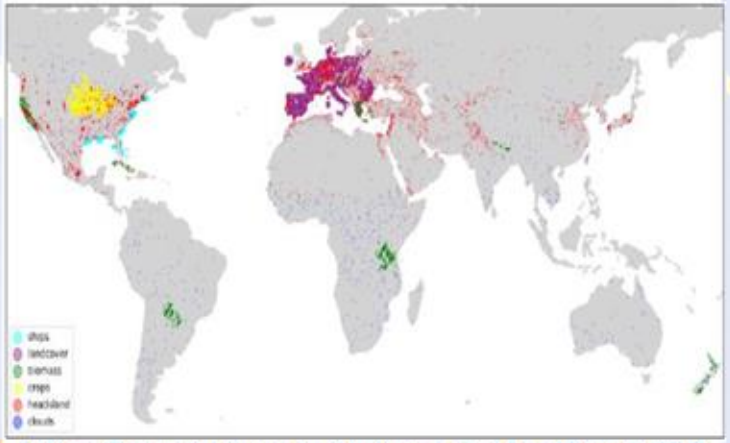
Embed2Scale Datasets

SSL4EO-S12 v1.1
an update with bugs fixed, served in >2GB of cloud-computing-ready ZARR format with integration into *TerraTorch*



https://github.com/DLR-MF-DAS/SSL4EO-S12-v1.1

SSL4EO-S12-downstream
200GB of SSL4EO-S12 data labeled for a diverse set of classification and regression tasks relevant to WP4



https://huggingface.co/datasets/embed2scale/SSL4EO-S12-downstream

SSL4Eco
200GB of SSL4EO-S12 data labeled for a diverse set of classification and regression tasks relevant to WP4

https://github.com/PlekhanovaElena/ssl4eco

Open-Source Model Zoo

GeoFE (CVPR data challenge winner)	https://github.com/torchgeo/torchgeo/pull/2868
FMFuse (CVPR data challenge runner-up)	https://github.com/KerekesDavid/embed2scale-solution
Temp-MOSAIKS (CVPR data challenge)	https://github.com/isaaccorley/temporal-mosaiks
SeCo-Eco (SSL4Eco top model)	https://github.com/PlekhanovaElena/ssl4eco
GSR4B (biomass super-resolution)	https://github.com/kaankaramanofficial/GSR4B
GSQ4EO (in-house ViT tokenizer)	https://github.com/embed2scale/GSQ4EO
S2-JEPA (in-house, I-JEPA for Sentinel-2)	*to be open-sourced*
TerraCodec (in-house timeseries compression)	*to be open-sourced*

Crop Stress and Yield Prediction

- Use county-level mean embeddings for the yield prediction (regression task).
- Baseline model used multimodality, so unfair comparison with Clay and Prithvi.
- However, we see comparable performance with AlphaEarth embeddings.