Climate Resilience through
Earth System Foundation Models

Carlos Gomes, Thomas Brunschwiler
IBM Research - Europe

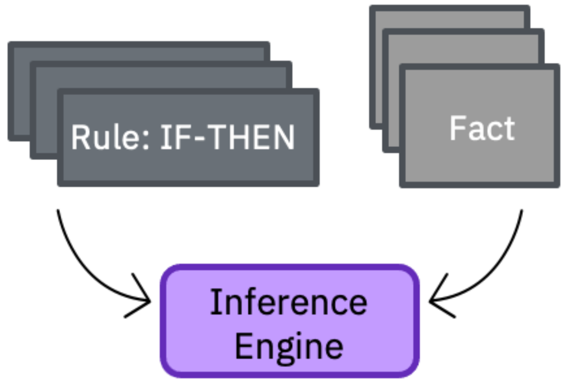
IBM

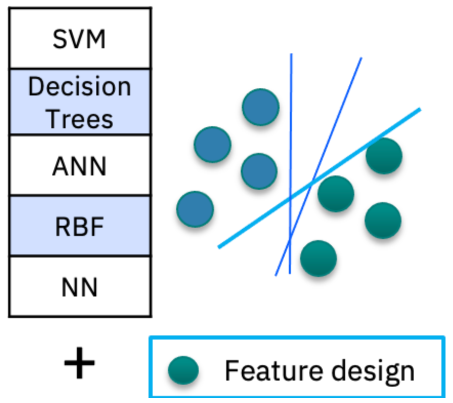
Foundational Technology Shift in AI Driving “Platform Dynamics”

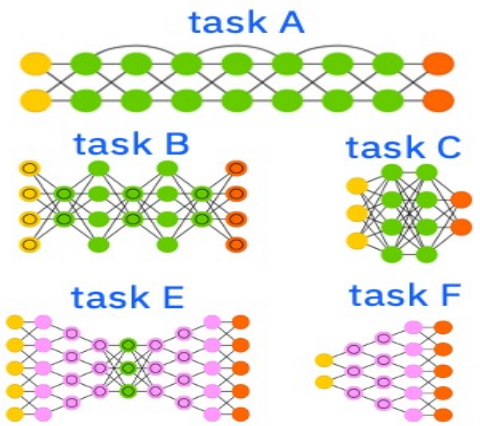
1980s | Big data

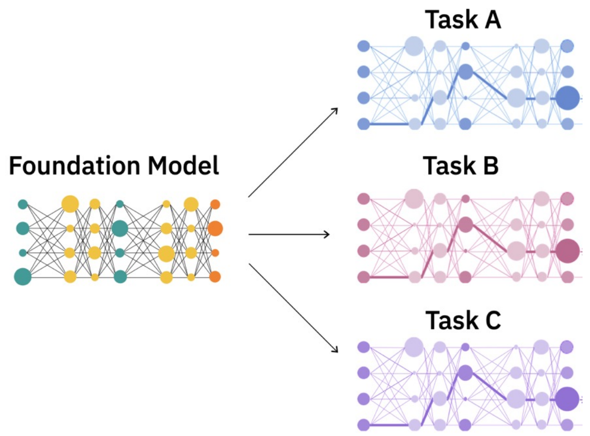
1980s to ~2010 | Massive labeled data + Compute

2017+ | **Self-supervision at scale** + Massive unlabeled data + Compute









Expert Systems
- Manually-crafted symbolic representations and rules
- No use of data and brittle

Machine Learning
- Less brittle but labor intensive
- Demanding data prep and feature engineering

Deep Learning
- Automatically learn if you have enough labeled data
- Enterprise adoption limited by availability of labeled data

Foundation Models
- Learn instrinsic structure from lots of data, *no need for labels*
- Quickly adopt to enterprise tasks using limited available labeled data

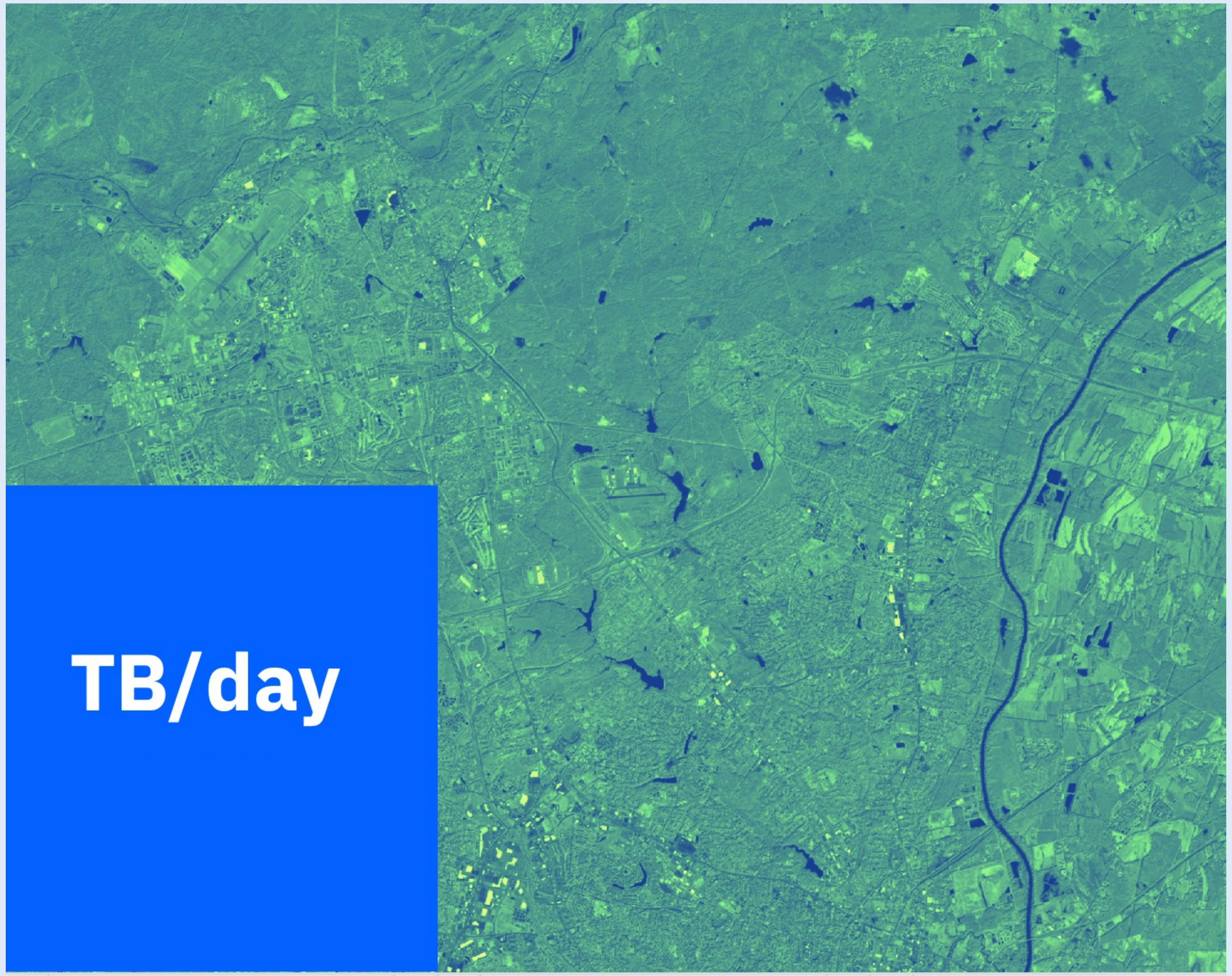
Outline

- Earth Observation FM & Applications
- Weather & Climate FMs vs. Emulators
- Climate Resilience through Earth System FMs
- Outlook: Embedding Engineering

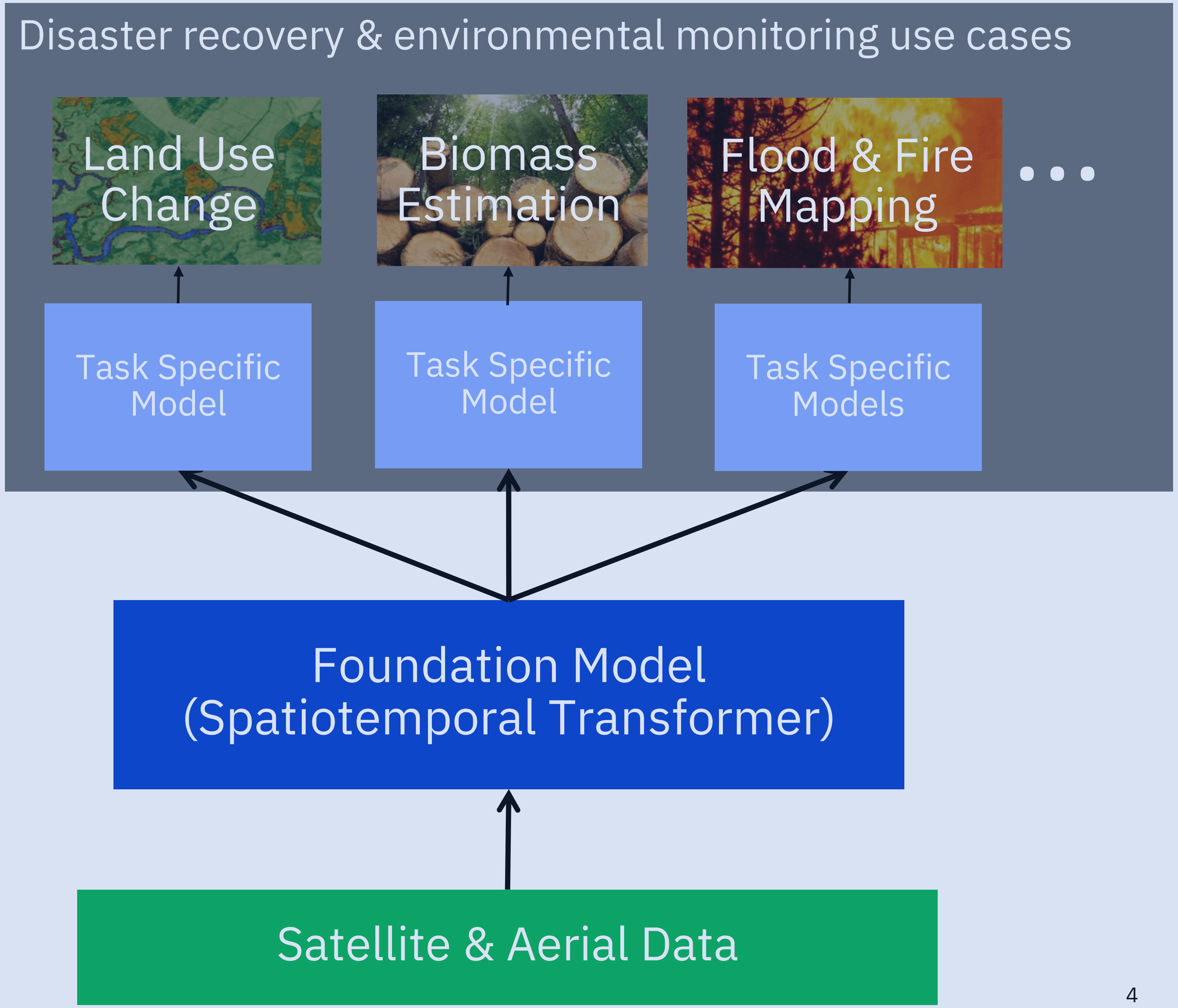
Earth Observation Foundational Models in collaboration with NASA

Satellite imagery

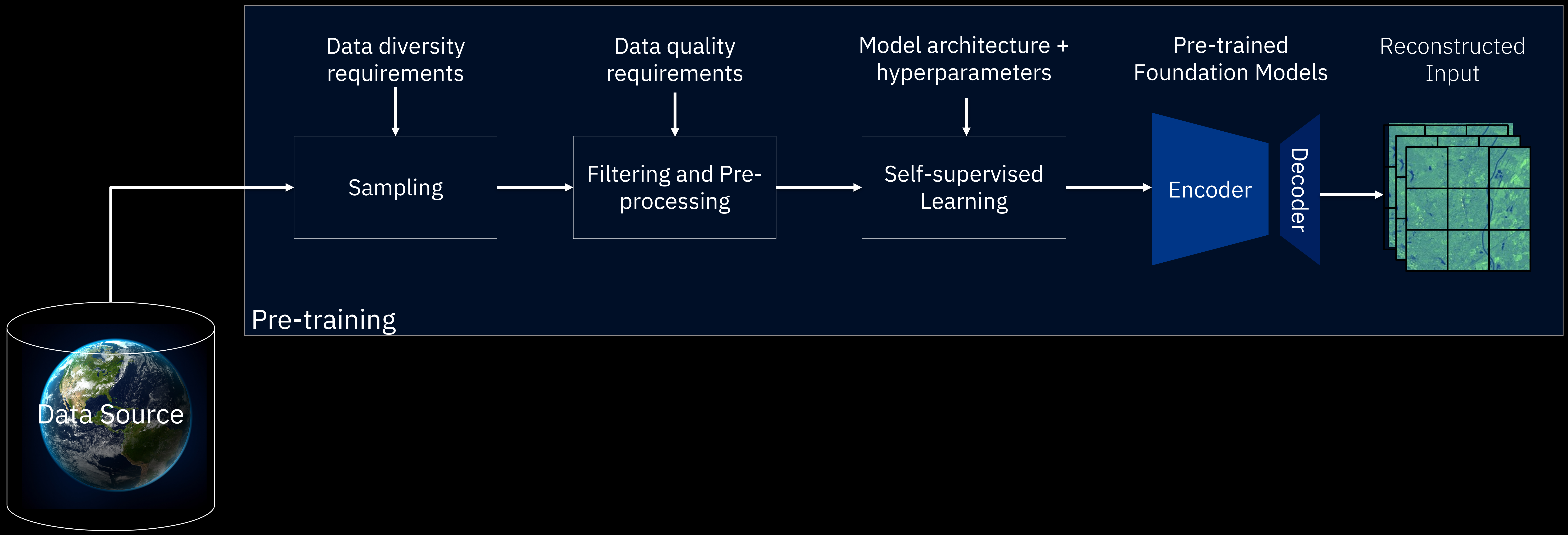
- Multimodal - multiple satellites and bands
- Temporal – regular updates



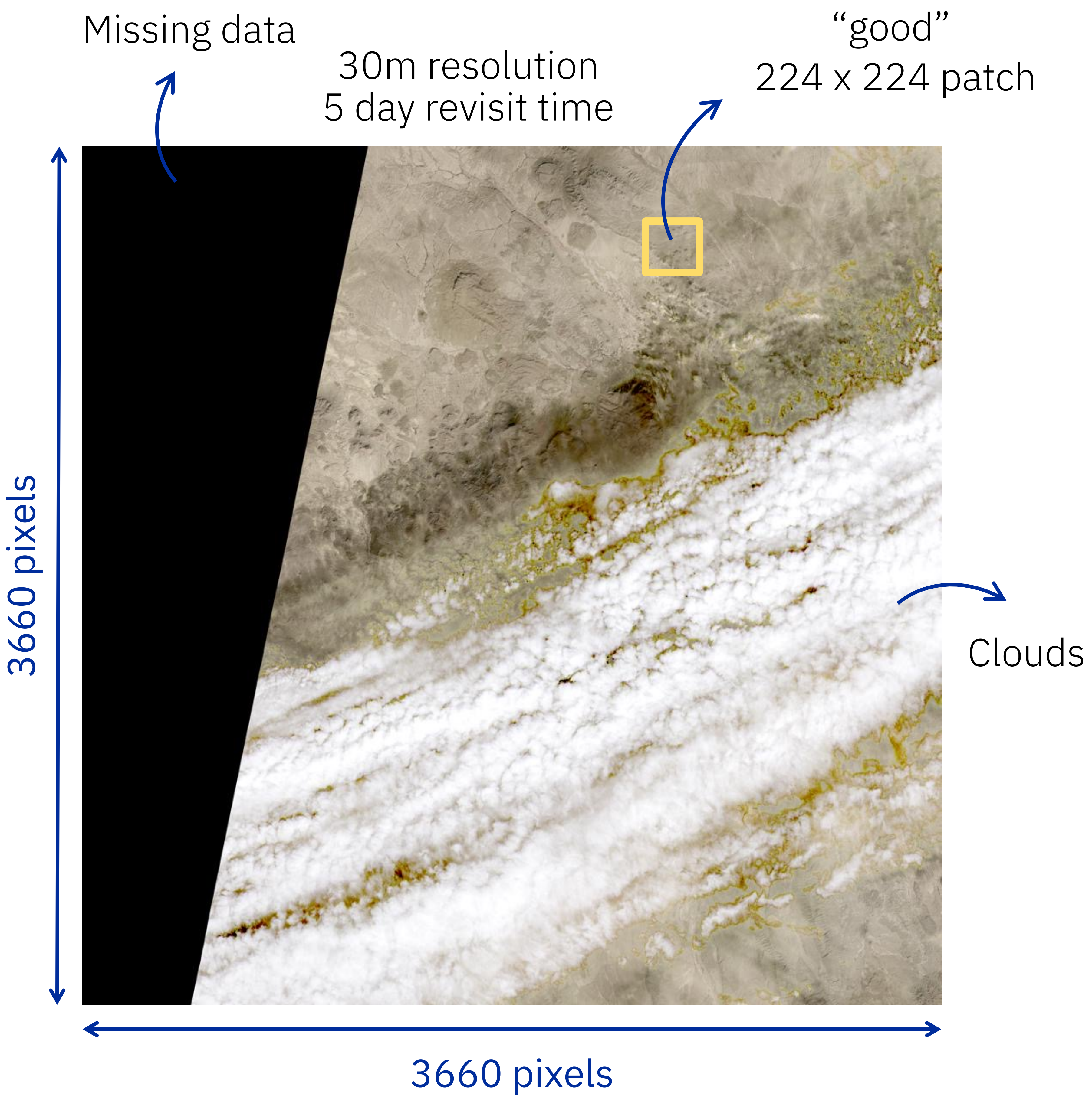
Geospatial Foundation Model



Foundation Models – Pre-training

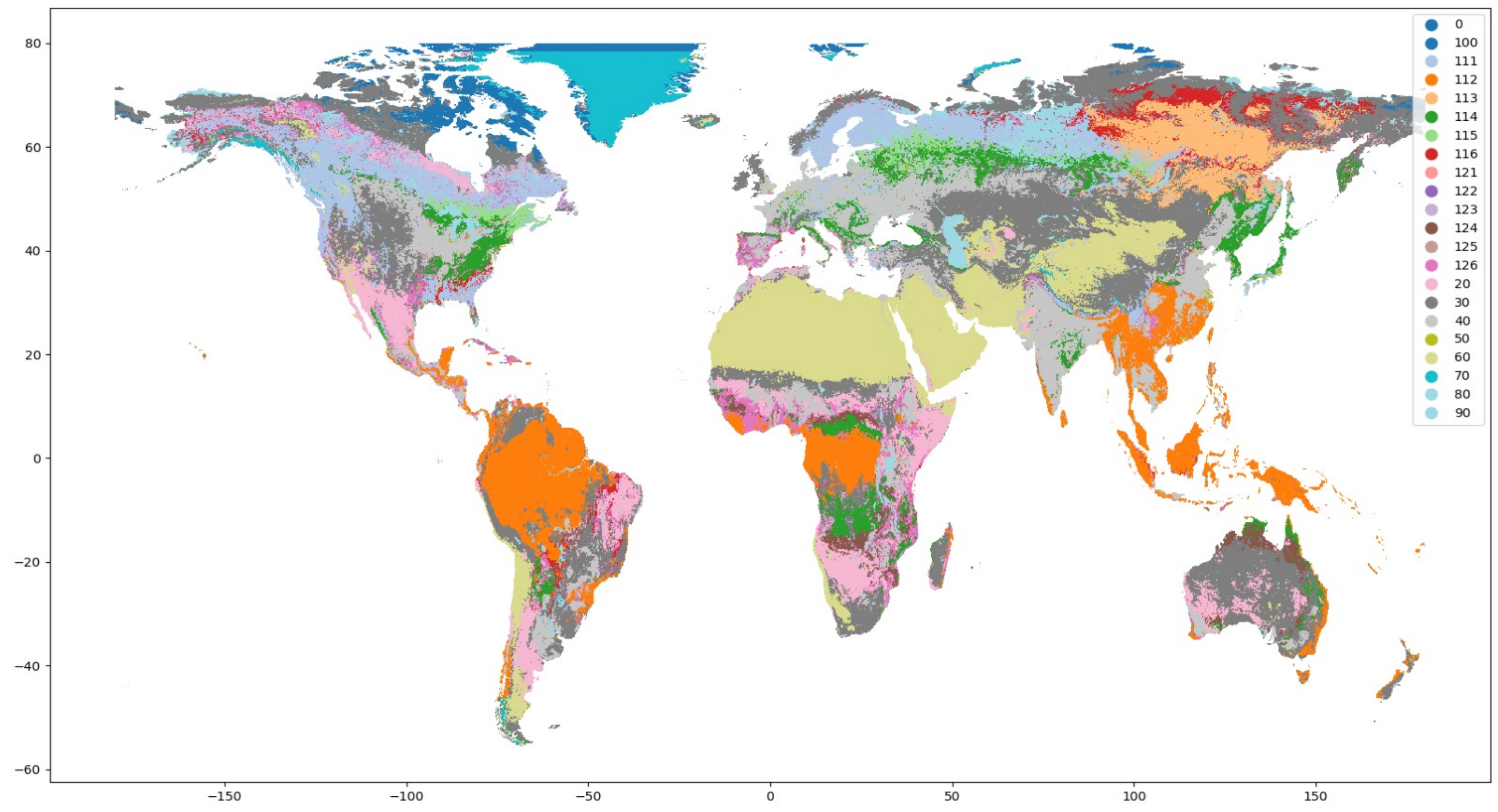


Harmonized Landsat-Sentinel Dataset Sampling



Diversity in pre-training data

- sampling based on geospatial statistics



Model architecture

MAE → Masked AutoEncoder

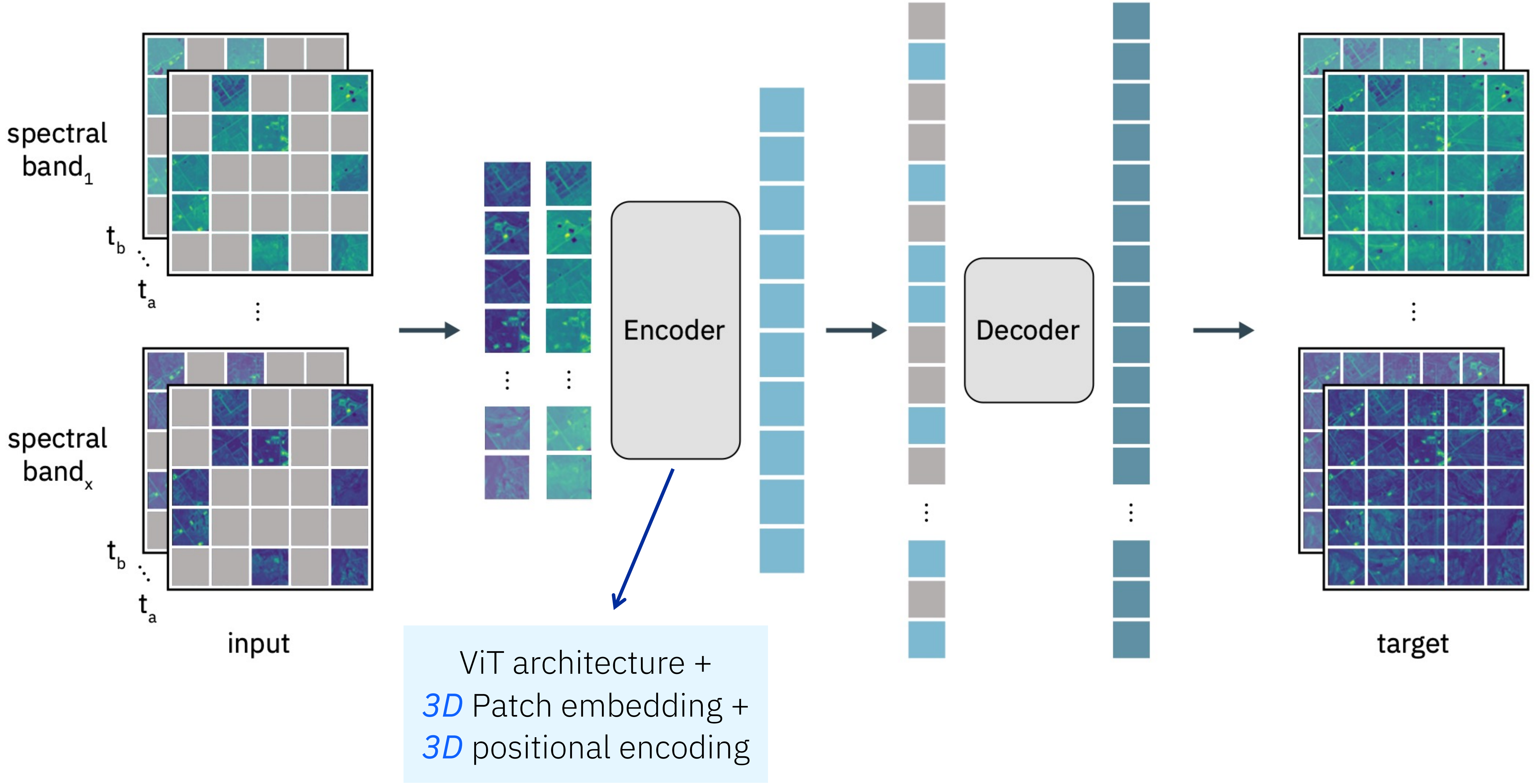
– Pre-training task: reconstruct *masked* patches → target = original data.

– MSE loss on *masked* patches.

Encoder → Vision transformer (*ViT*) for multispectral *3D data*.

– 3D patch embeddings
– 3D positional encoding

Decoder → Transformer blocks + linear projection layer to match the target patch size.



IBM NASA Geospatial

Community

Research interests

Geospatial foundation models using HLS2 data

Team members 38

Organization Card

About org cards

NASA and IBM have teamed up to create an AI Foundation Model for Earth Observations, using large-scale satellite and remote sensing data, including the Harmonized Landsat and Sentinel-2 (HLS) data. By embracing the principles of open AI and open science, both organizations are actively contributing to the global mission of promoting knowledge sharing and accelerating innovations in addressing critical environmental challenges. With Hugging Face's platform, they simplify geospatial model training and deployment, making it accessible for open science users, startups, and enterprises on multi-cloud AI platforms like watsonx. Additionally, Hugging Face enables easy sharing of the pipelines of the model family, which our team calls Prithvi, within the community, fostering global collaboration and engagement.

Watch Prithvi end to end demo

More information: NASA Blog post, NASA Veda system, IBM Press/Blog post, EIS, Code

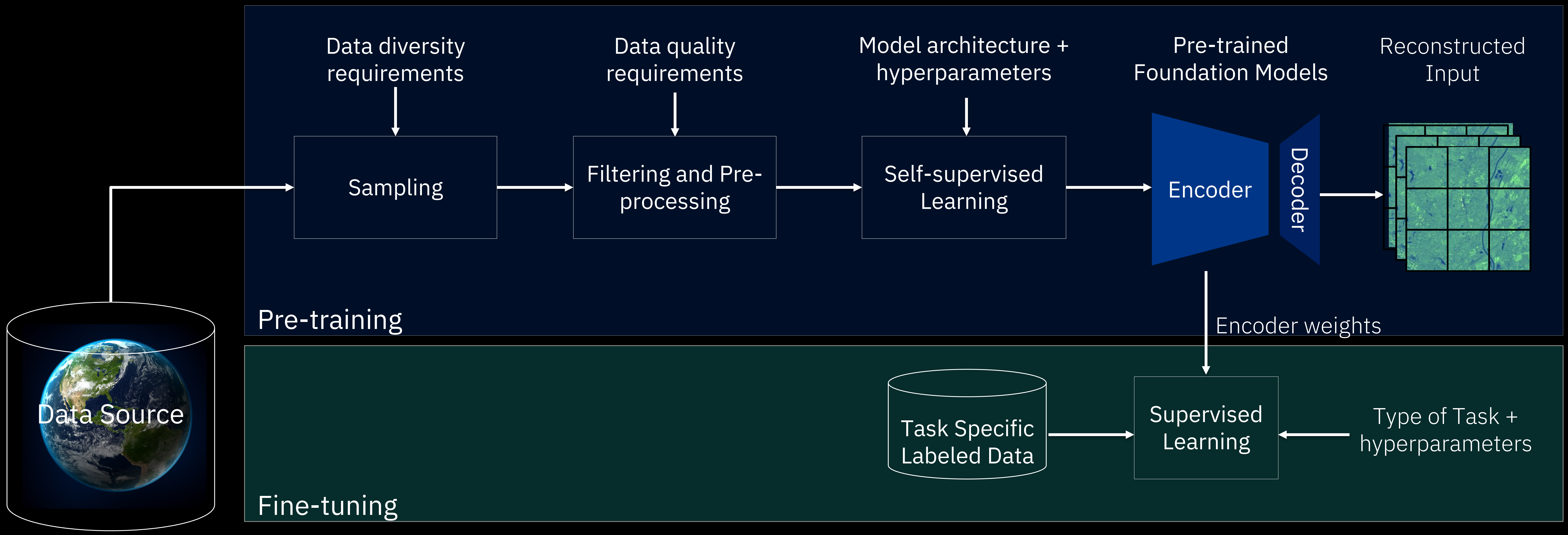
Spaces 4

Sort: Recently Updated

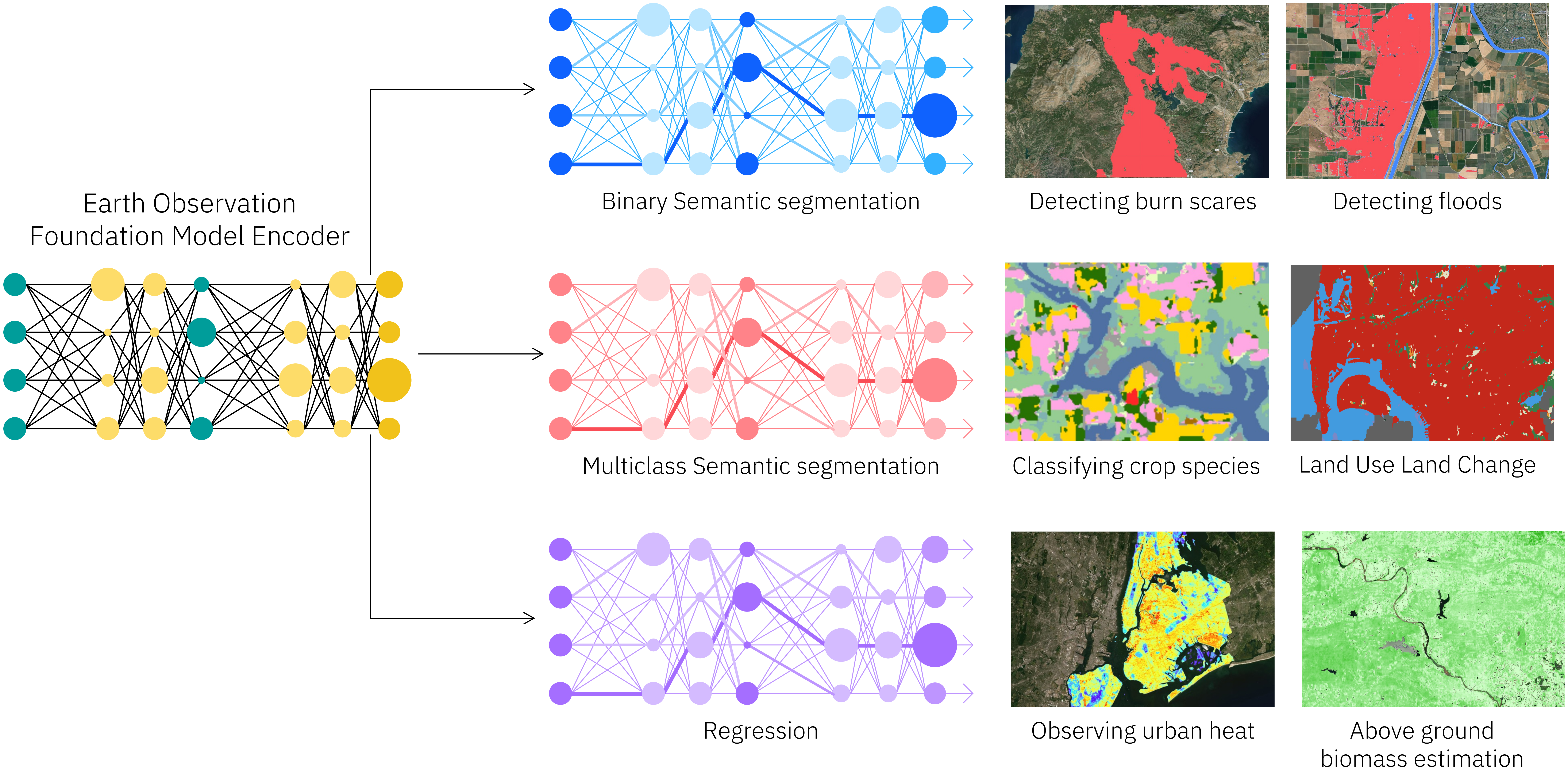
18

14

Foundation Models – Fine-tuning

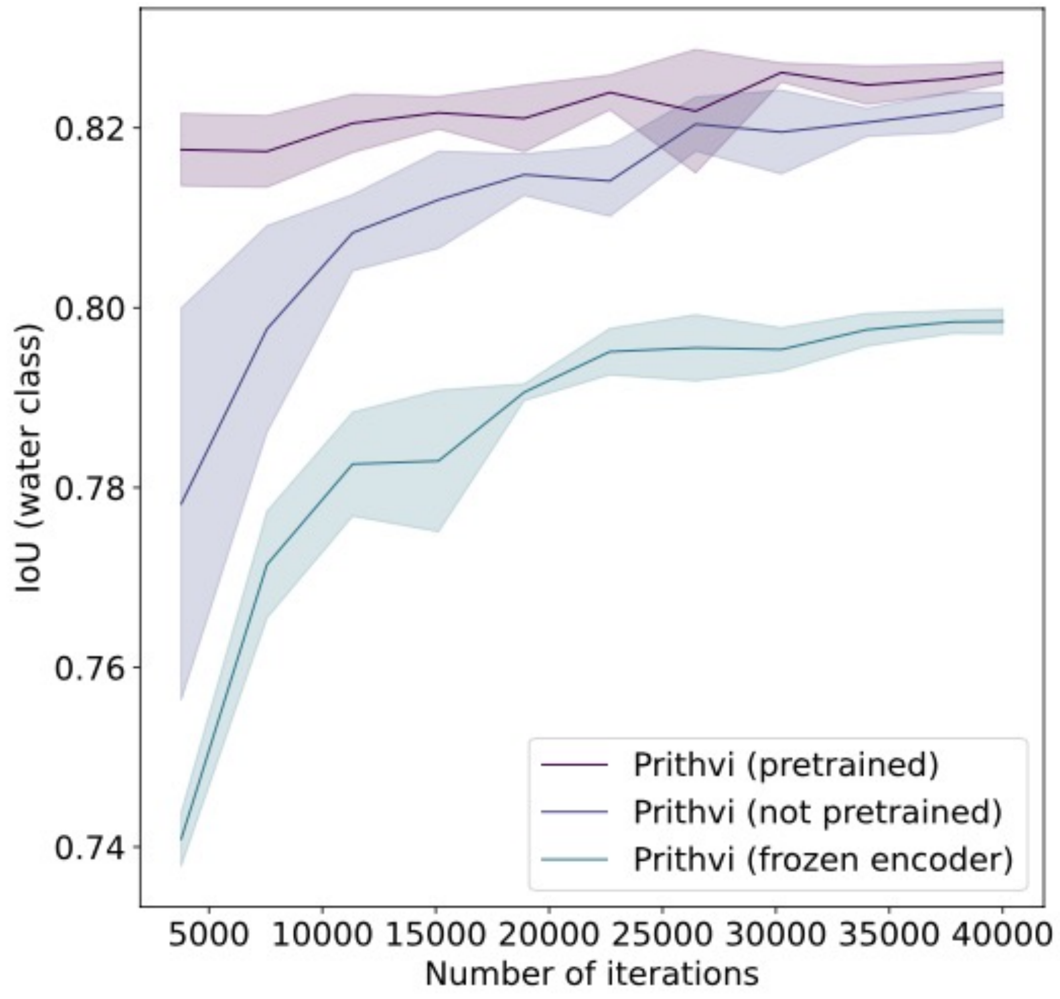


One trained Geospatial Foundation Model for many tasks

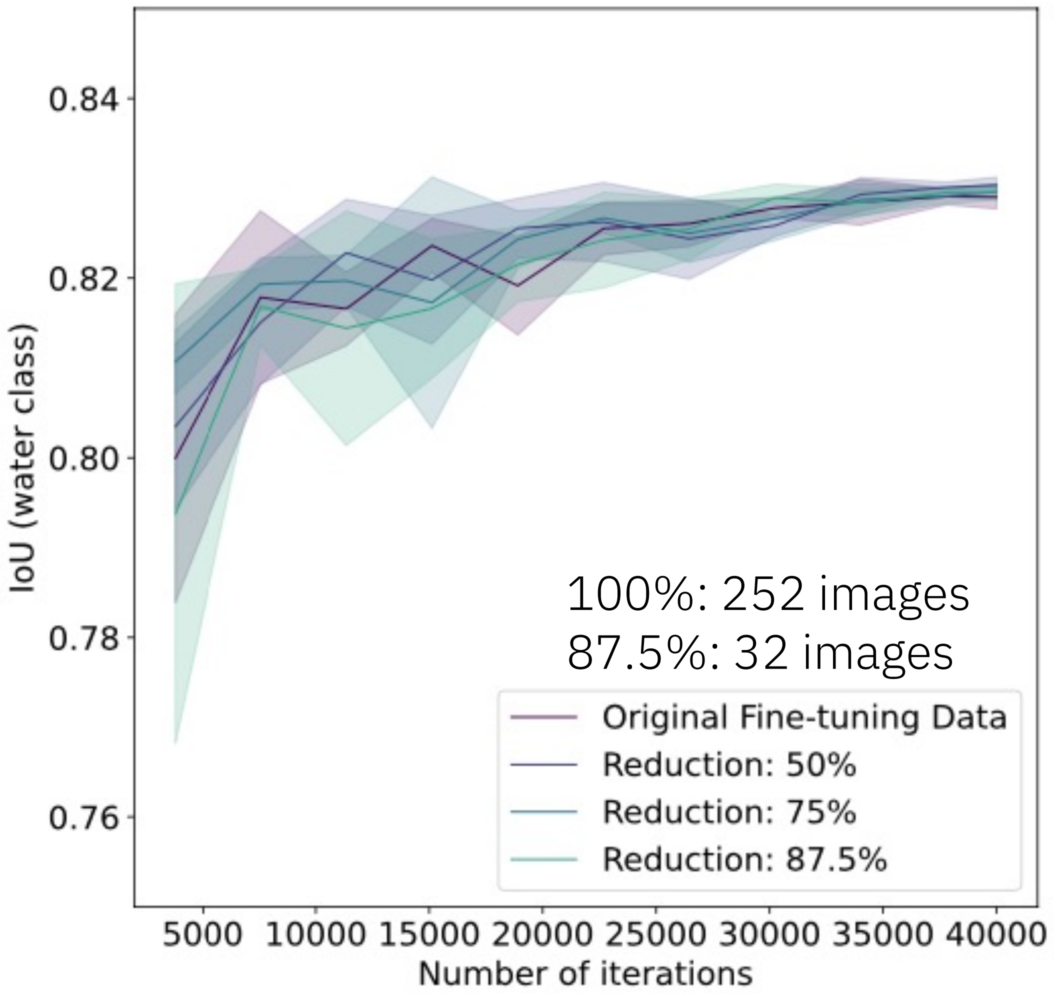


EO-FM KPI's on Flood Mapping - fine-tuning analysis

Encoder Weights



Data Efficiency

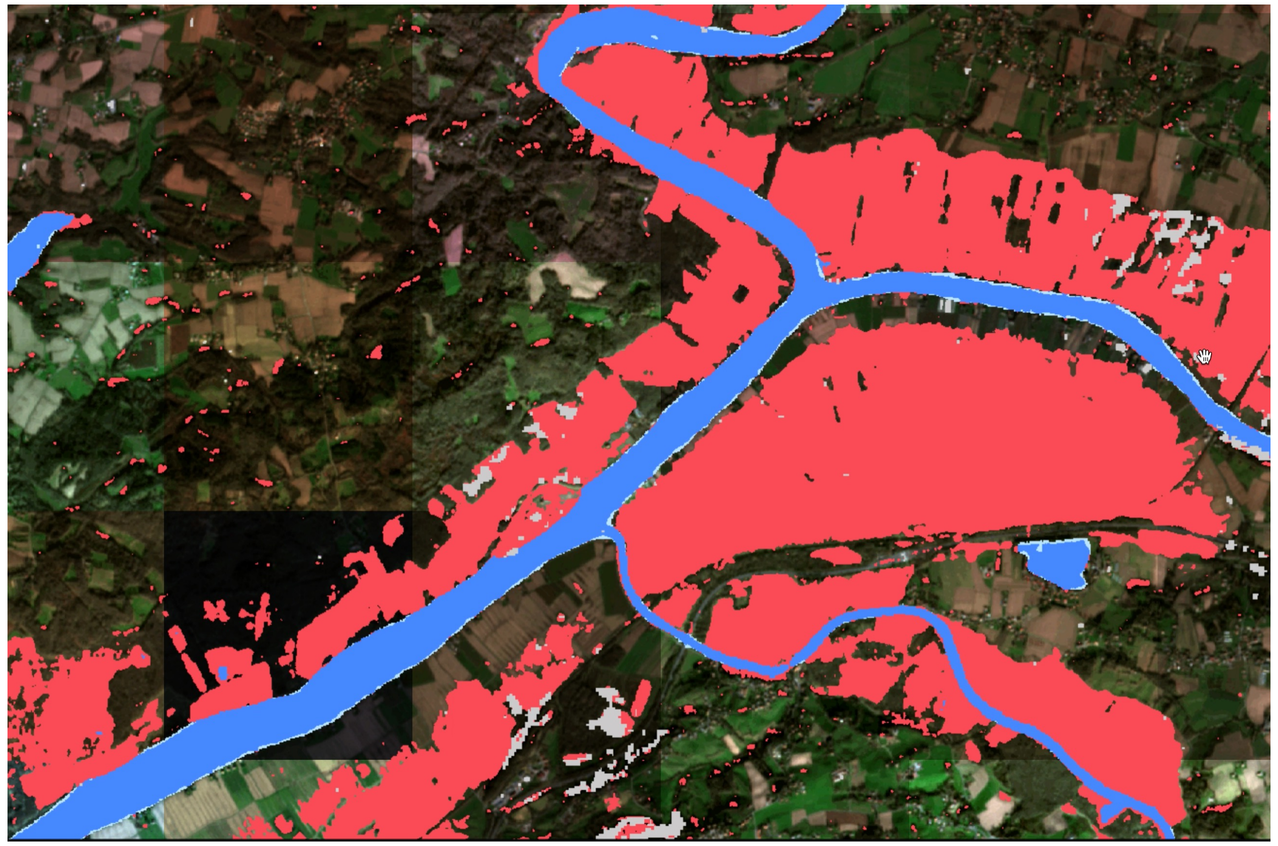


	IoU (water)	F1 (water)	mIoU (both classes)	mF1-score (both classes)	mAcc (both classes)
Baseline [55]	24.21	–	–	–	–
ViT-base [19]	67.58	80.65	81.06	88.92	88.82
Swin [60]	79.43	88.54	87.48	93.13	90.63
Swin† [60]	80.58	89.24	87.98	93.44	92.02
After 50 epochs					
Prithvi (not pretrained)	80.67	89.30	88.76	93.85	94.79
Prithvi (pretrained)	81.26	89.66	89.10	94.05	**95.07**
After 500 epochs					
Prithvi (not pretrained)	82.97	90.69	90.14	94.66	94.82
Prithvi (pretrained)	**82.99**	**90.71**	**90.16**	**94.68**	94.60

Key differentiations

- Pretrained Prithvi accelerates fine-tuning
- Outperforms state-of-the-art 15%
- Performs well in few-shot learning mode
- Generalizes well across global regions

blue: permanent waters, red: flooded area



QuickTime Player File Edit View Window Help

https://geostudio-ui-nasageospatial-dev.cash.sl.cloud9.ibm.com

IBM watsonx.ai for Geospatial

Welcome Thomas Brunschwiler!

An internal service from IBM Research, prototyping generative AI technology and tooling for watsonx.ai

Inferencing Lab

Experiment with foundation models

Open inferencing lab →

Fine-tuning Studio

Fine-tune your foundation models with labeled data

Tune a model →

Documentation

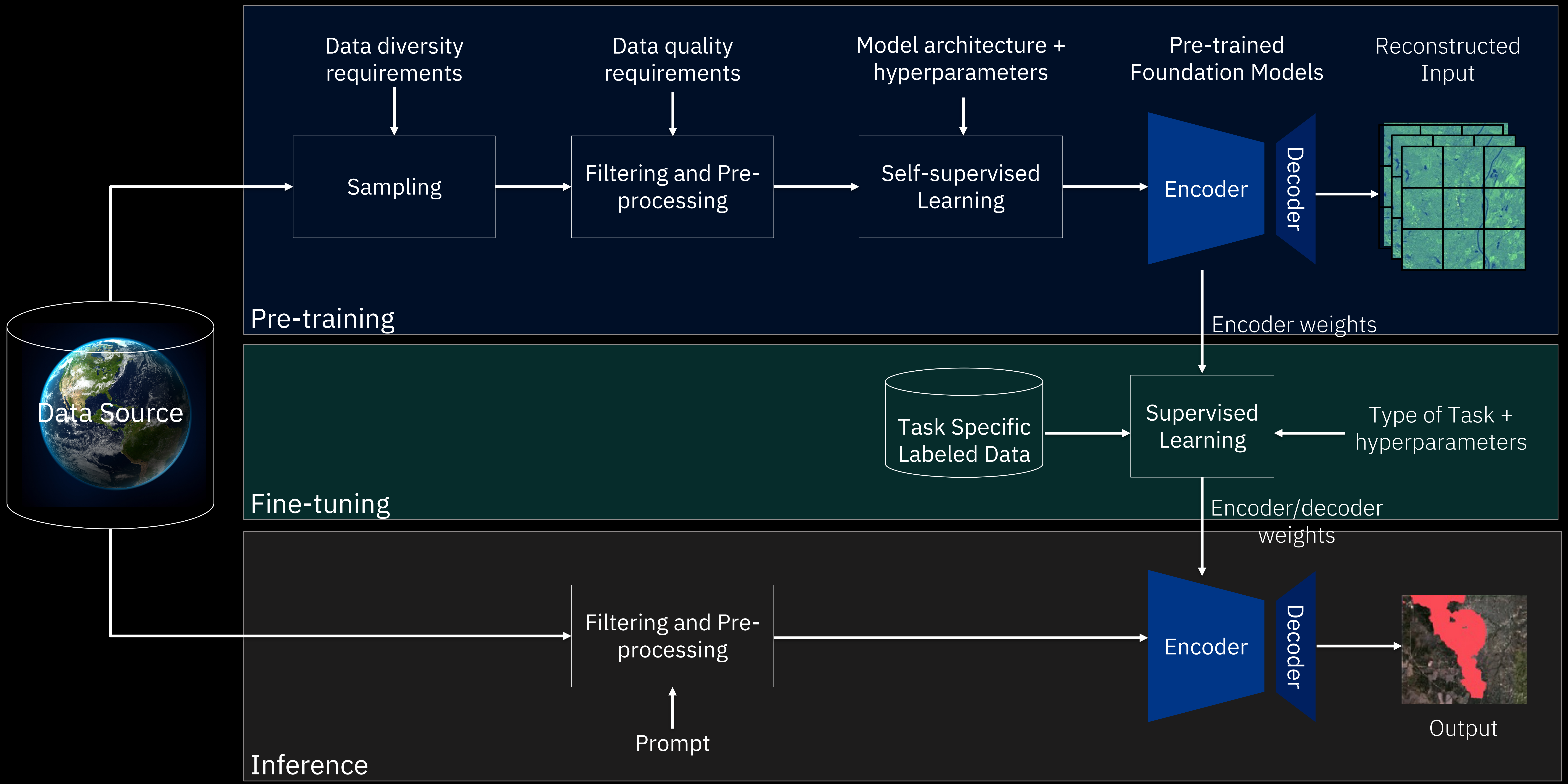
Access SDKs, model cards, and developer resources

API Key

*****

Read documentation →

Foundation Models - Inference

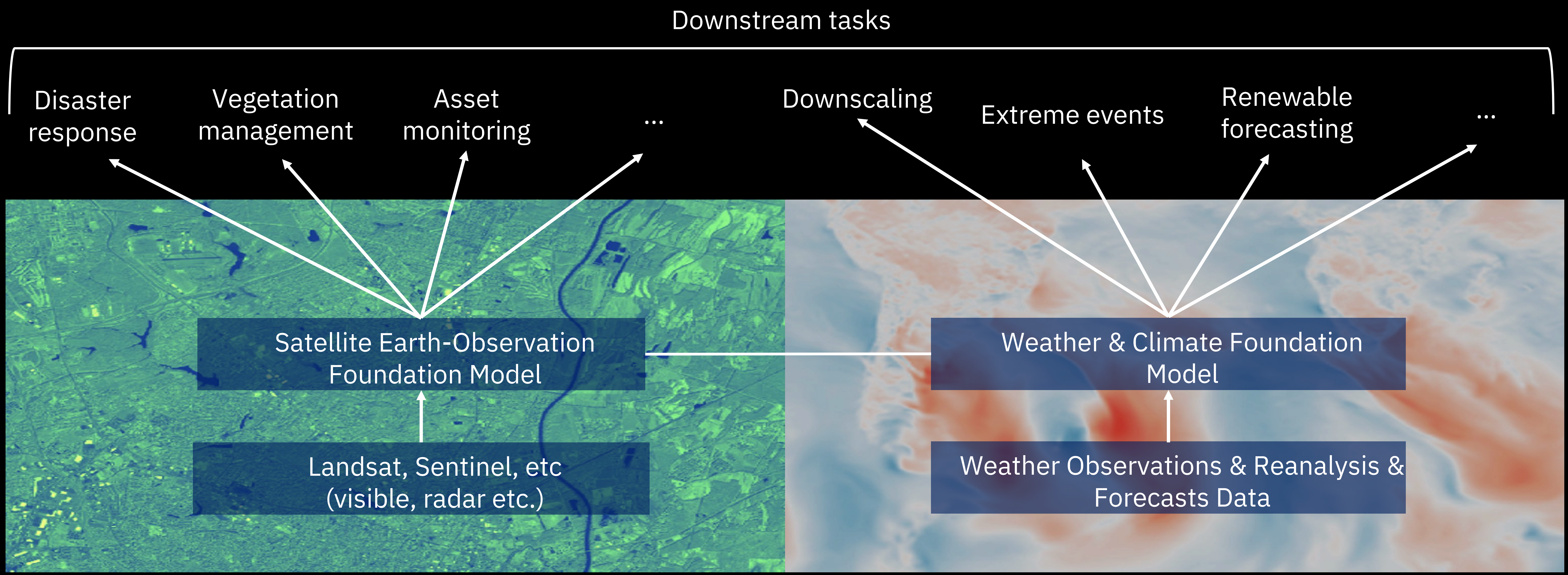


Hands-On exploration

https://github.com/NASA-IMPACT/hls-foundation-os/blob/main/exploration.ipynb

Weather & Climate Foundation Model

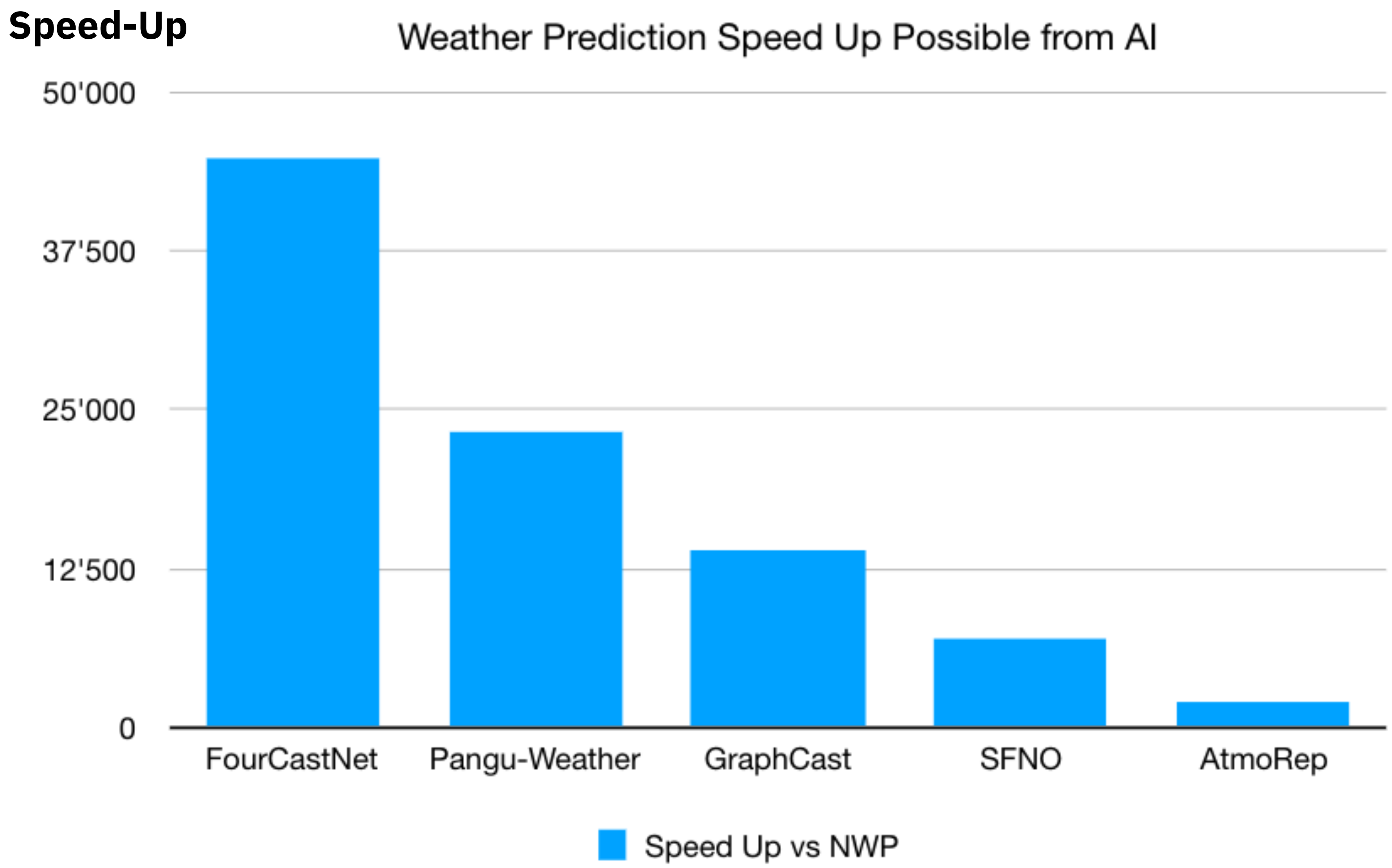
Earth System Foundation Models



Development Partners

NASA
horizon europe
The AI Alliance
Environment and Climate Change Canada
Environnement et Changement climatique Canada
HUGGING FACE

AI Model Forecasting – orders of magnitude speed-up



(Latency Node-Seconds: IFS / FCN(18km) Ratio) "IFS takes 984000 sec (273.3hrs) to compute a 24hr forecast for a 100-member ensemble whereas FCN takes 7 sec, 30km and 22 sec, 18km res, on a single 4GPU A100 node. In comparison, the IFS model needs 3060 nodes for such a forecast". Pg17-18

FCN 1 member = 7sec/100*4 = 0.28sec on A100 whereas Pangu 1 member = 1.4s*(120TFlops/312TFlops) = 0.53s on A100, so **Pangu for 100member on 4A100node = 13.46sec** i.e. 42.30 sec at 18km (extrapolated), 98400/42.3 = 23262 Latency Node-Seconds: IFS / Pangu (18km) Ratio

10 day forecast ~ 1 min on TPUv4 ~ 1.5 min on A100, whereas NWP HRES takes 21120 minutes (352nodesx60min), GraphCast provides speed up of 14'080

1h clock*1000 nodes/12.8m A6000 node *1.5 (if on A100)

1000 dual-socket CPU nodes (Bauer et al., 2020). With the caveat of differing hardware, this corresponds to a speedup of close to 5,000x" https://arxiv.org/abs/2306.03838

10sx8x100/16~500s for a 24h 100member i.e. 500/22 ~ 22.7 x slower than FourCastNet

"10s for a global 3-hour forecast on 1 node (4A100)" pg28 "16 ensemble members" pg 14

All trained with ERA5 (re-analysis data from ECMWF)

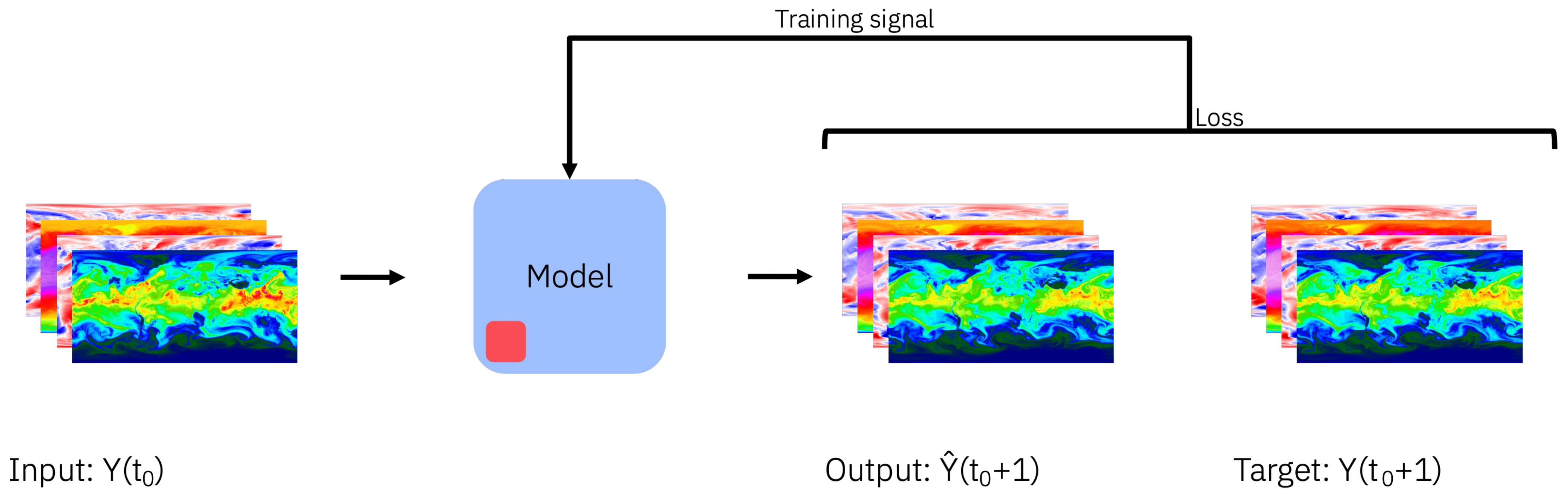
BUT does not include model **training resources**

-> motivates need for **Foundation Models**

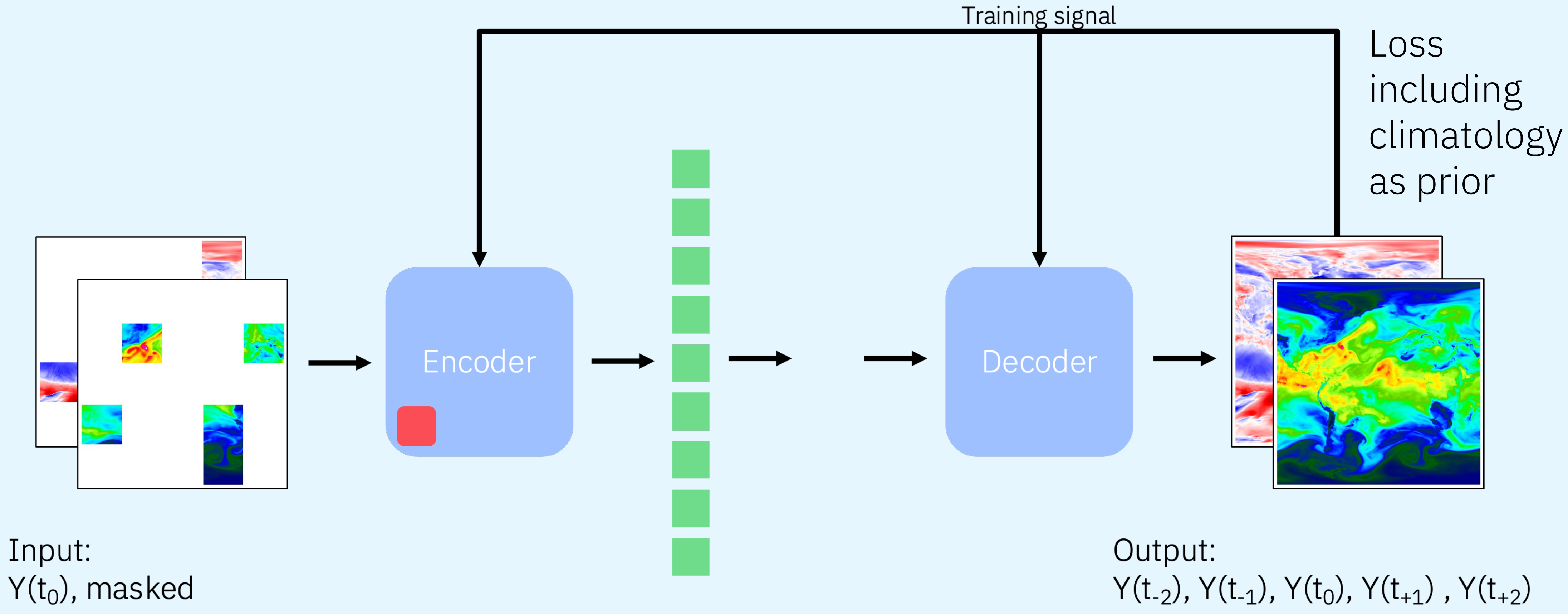
AI emulator	Training Resources	~ V100 GPU Hours
FourCastNet AFNO	16h x 64 A100 GPU	1'664
GraphCast	4wk, 32 Cloud TPUv4	147'456
Pangu	4 x 16days x 192 V100	294'912

From AI Forecast Emulators to Weather Foundation Models

AI-based forecast emulators are trained to propagate data in a fixed representation into the future.

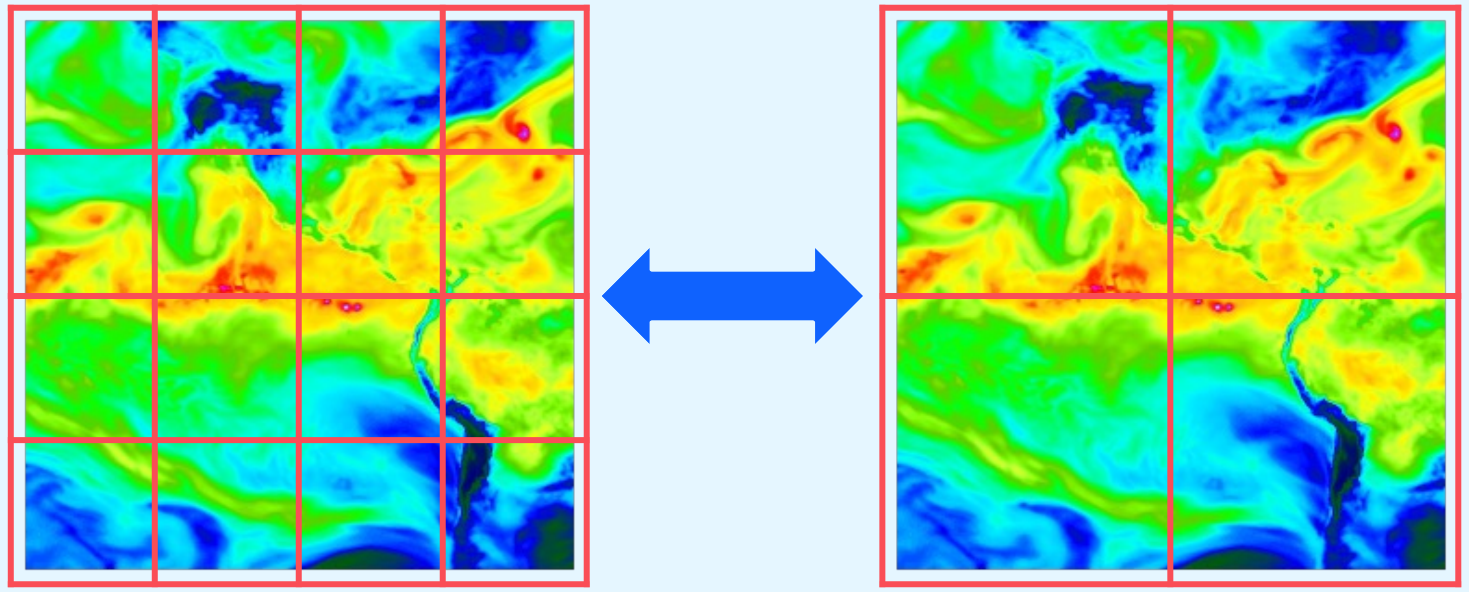


Foundation models are trained on a pretext task – e.g., reconstruction of t_{-n} to t_{+n} – to learn patterns in global weather that can be leveraged beyond forecasting.

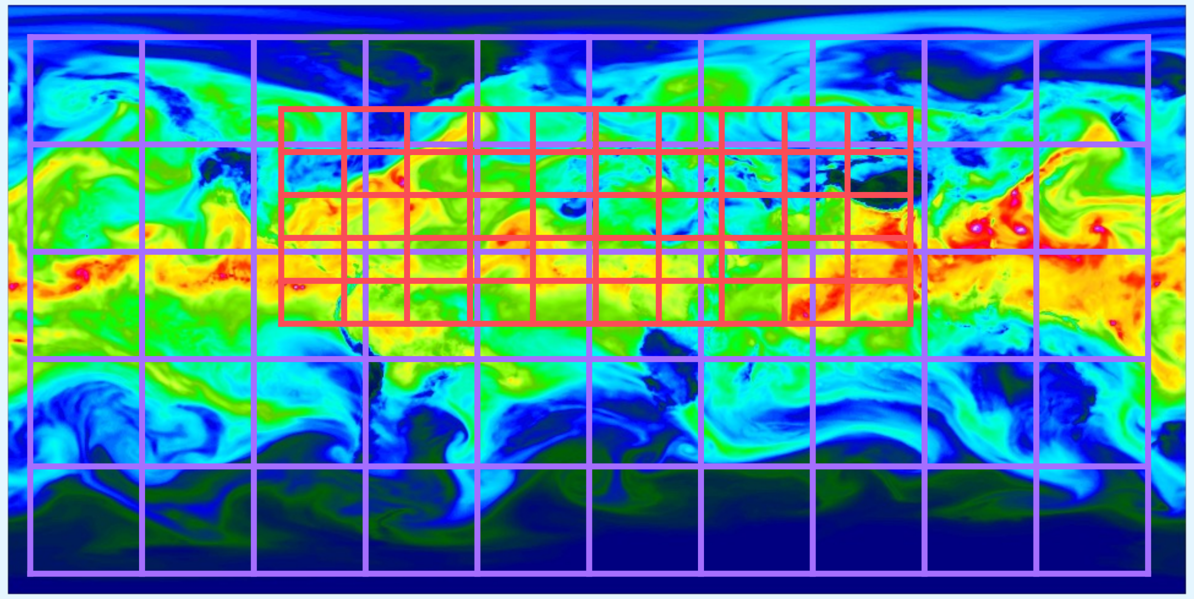


Benefits of our Transformer Architecture for Weather Modelling

Tokenization:
A pathway to multi-resolution models



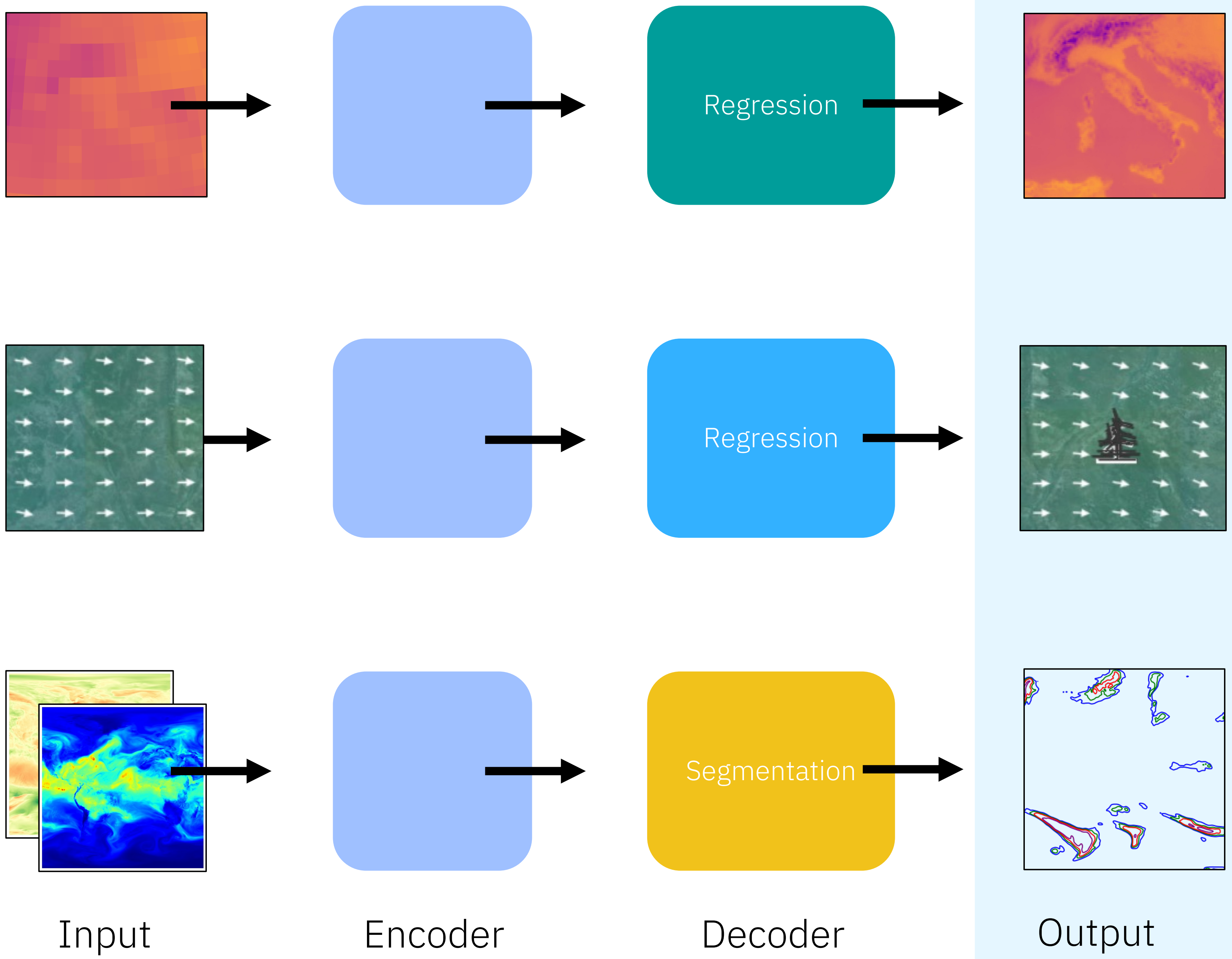
Flexible sequence length:
A pathway to multi-regional models



Tokenization + flexible sequence:
A pathway to integrating observations

Pretrained FMs facilitate Different Challenges

Once pre-trained encoder can be combined with various task-specific decoders to solve specific applications.

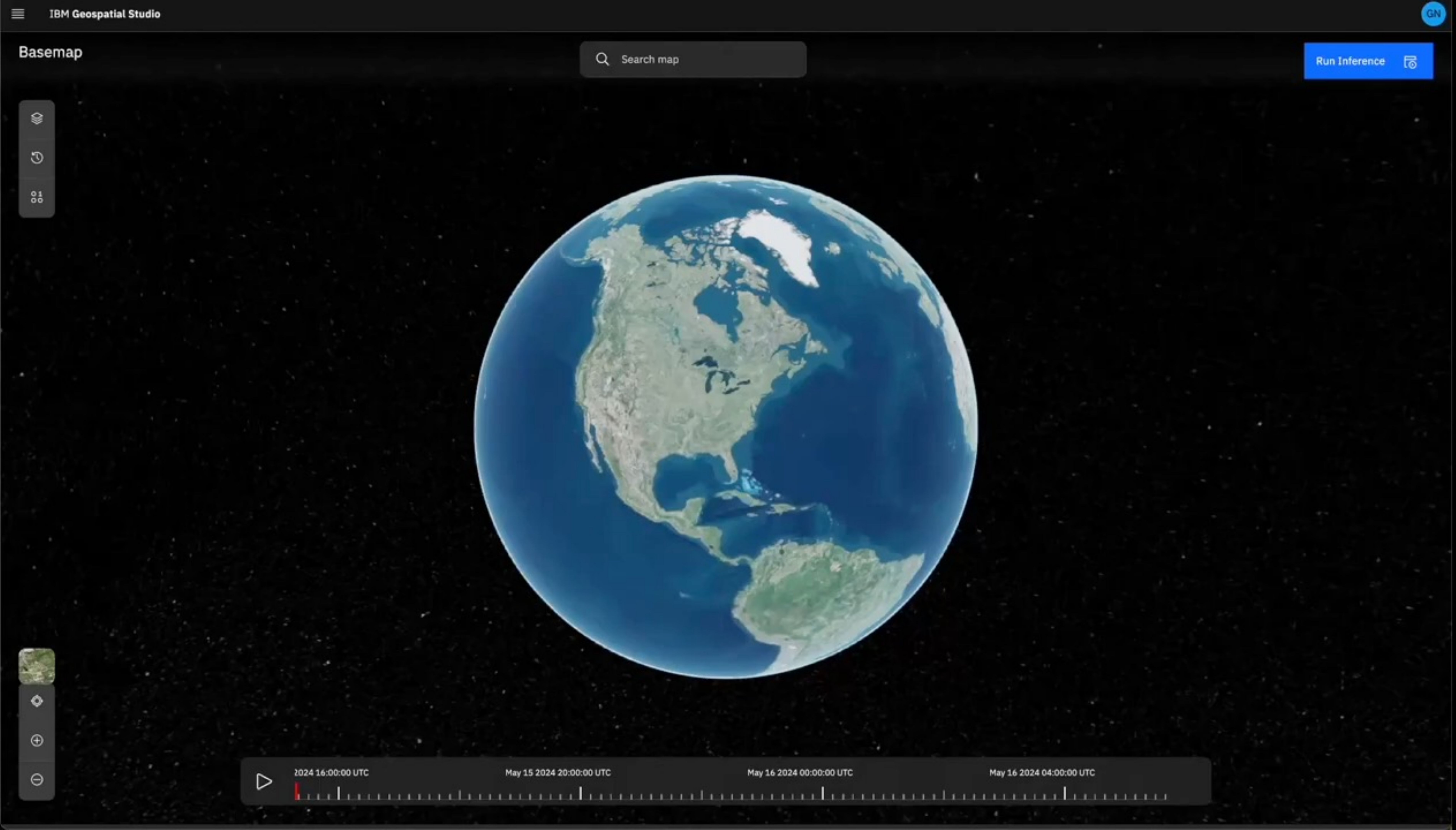


Climate Downscaling

Off-grid Wind Forecasting

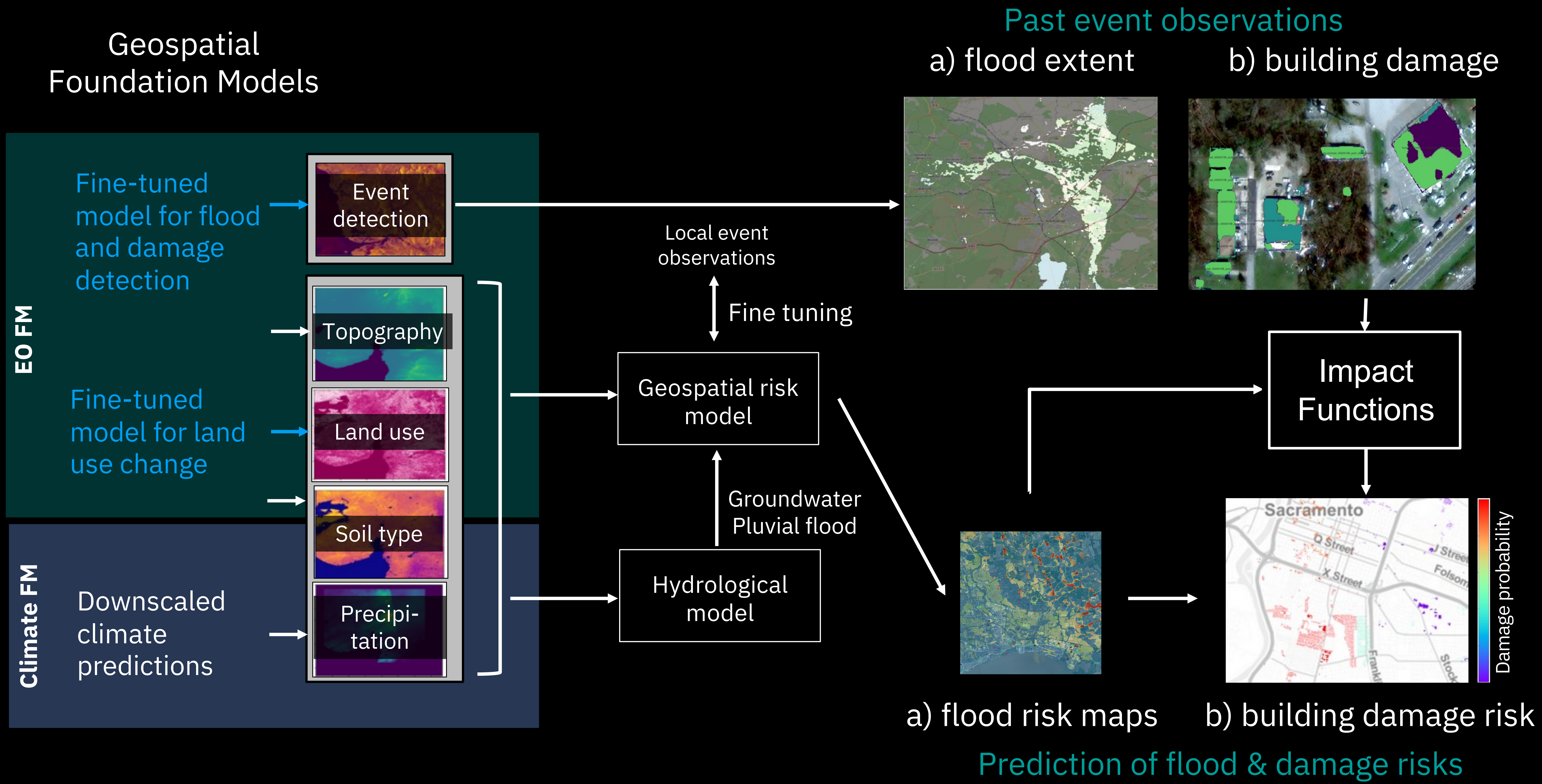
Extreme Event Detection

⋮


IBM Geospatial Studio
GN
Basemap
Search map
Run Inference
2024 16:00:00 UTC
May 15 2024 20:00:00 UTC
May 16 2024 00:00:00 UTC
May 16 2024 04:00:00 UTC

Climate Resilience through Earth System FMs

Potential Model Workflow for Hydrogeologic Risk



High Resolution Urban Heat Island Estimation

Satellite spectral and thermal bands (Landsat)

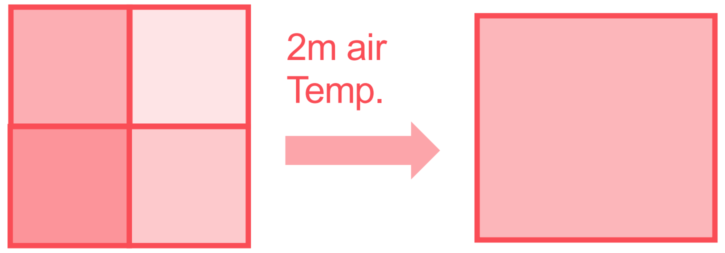
- Reflectance bands: Changes over weeks/months
- Thermal: Changes hourly
- 30m res., 16-day revisit (global)

Red
Green
Blue
SWIR
NIR
…

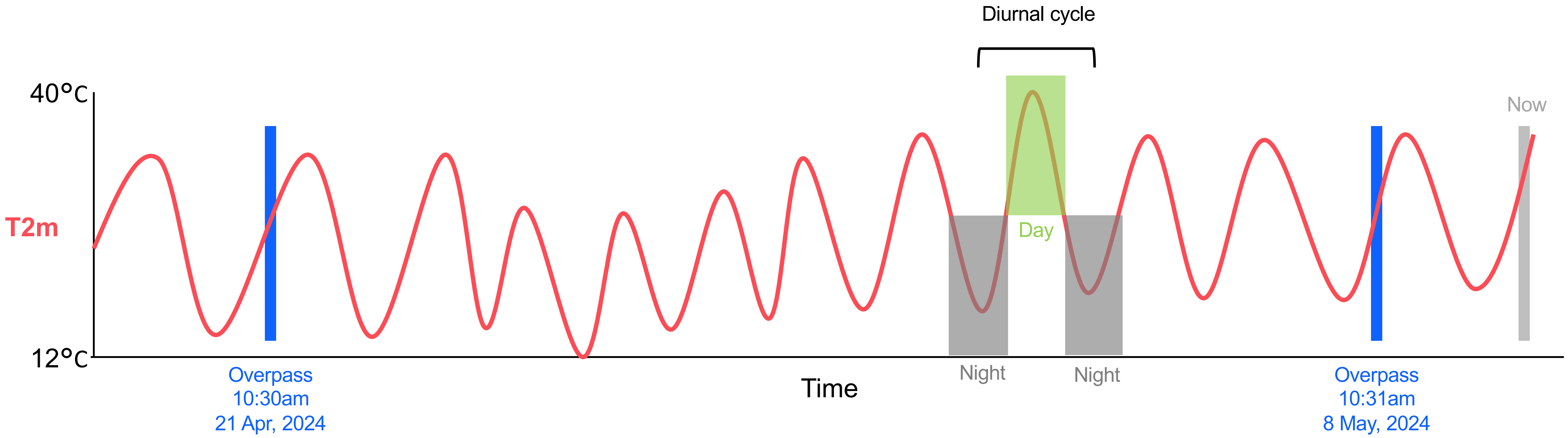
SLR
LST

Reanalysis/Climate variables (ERA5)

- T2m: Changes hourly
- Available hourly (global)
- 9km res.
- Different variables



Average
Regrid



High Resolution Urban Heat Island Estimation

Satellite spectral and thermal bands (Landsat)

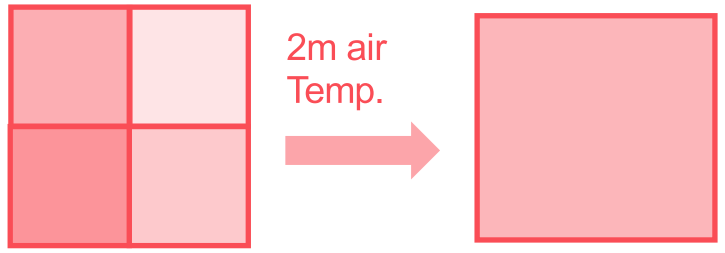
- Reflectance bands: Changes over weeks/months
- Thermal: Changes hourly
- 30m res., 16-day revisit (global)

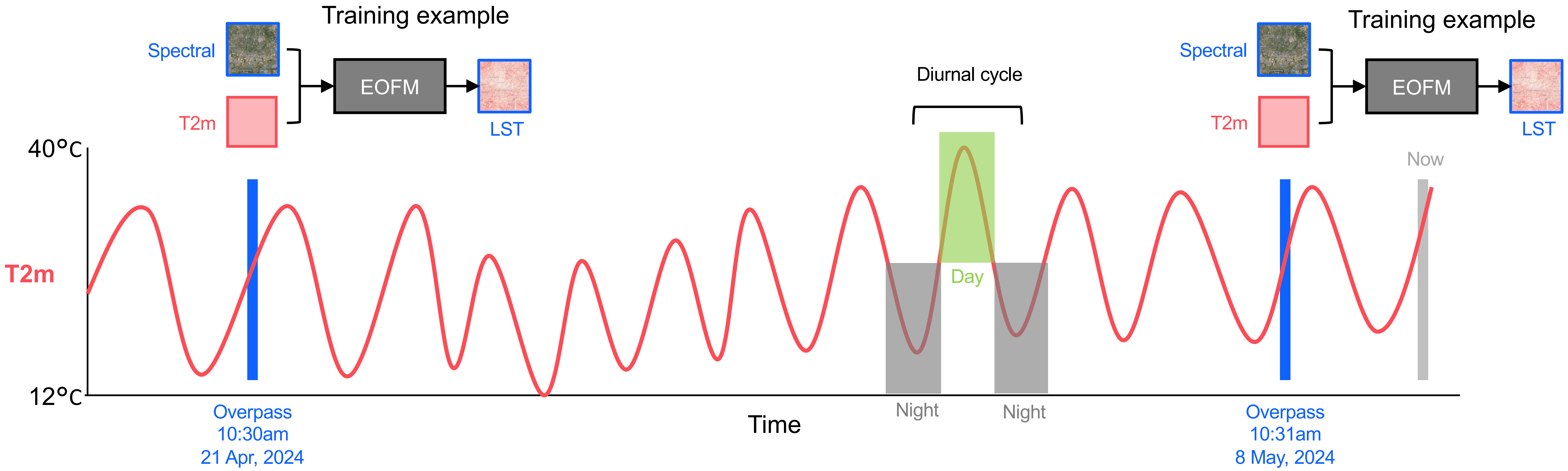
Red
Green
Blue
SWIR
NIR
…

SLR
LST

Reanalysis/Climate variables (ERA5)

- T2m: Changes hourly
- Available hourly (global)
- 9km res.
- Different variables





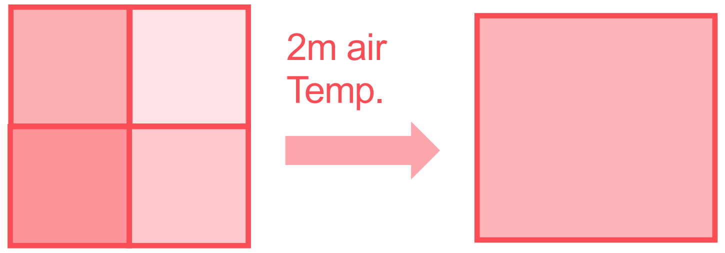
High Resolution Urban Heat Island Estimation

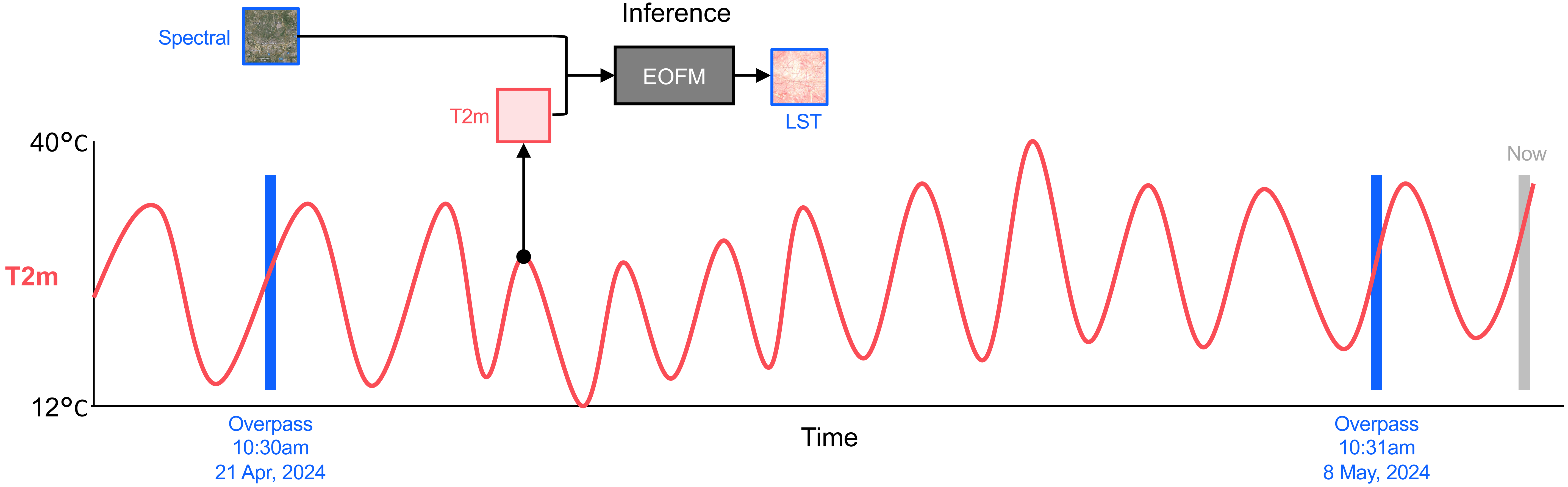
Satellite spectral and thermal bands (Landsat)

- Reflectance bands: Changes over weeks/months
- Thermal: Changes hourly
- 30m res., 16-day revisit (global)

Reanalysis/Climate variables (ERA5)

- T2m: Changes hourly
- Available hourly (global)
- 9km res.
- Different variables





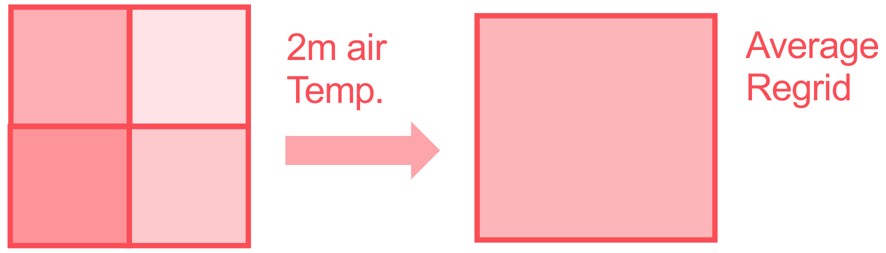
High Resolution Urban Heat Island Estimation

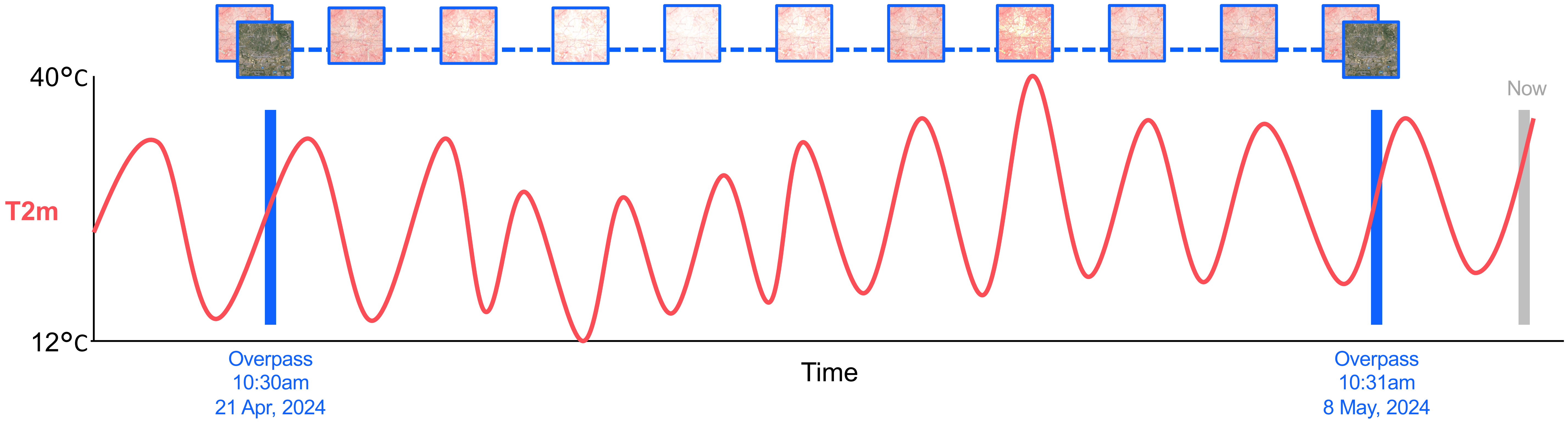
Satellite spectral and thermal bands (Landsat)

- Reflectance bands: Changes over weeks/months
- Thermal: Changes hourly
- 30m res., 16-day revisit (global)

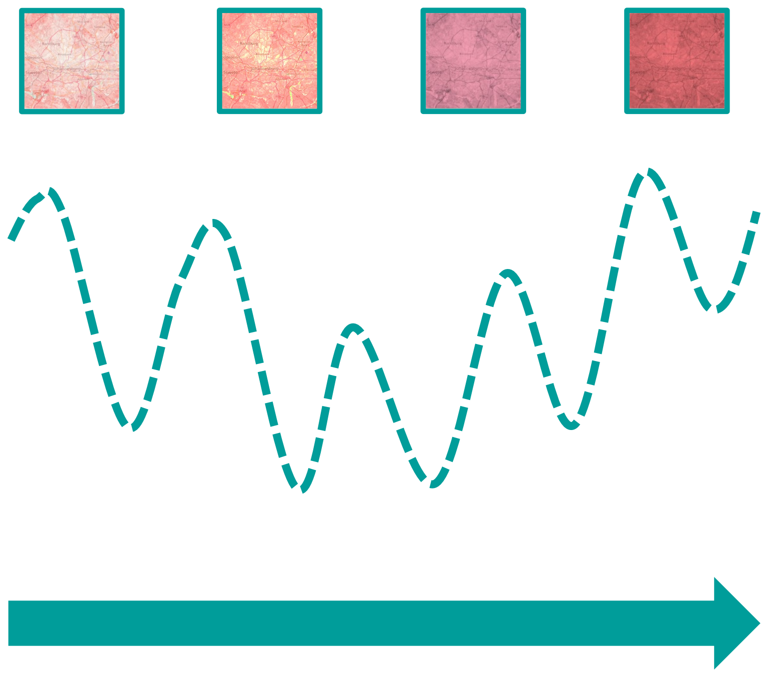
Reanalysis/Climate variables (ERA5)

- T2m: Changes hourly
- Available hourly (global)
- 9km res.
- Different variables





Forecasts & Projections



Weather (1-2 weeks)
S2S (2 weeks – 3 months)
Seasonal (3 – 6 months)
Climate (1 – 10+ years)

ECMWF
CMIP6


IBM Research – Zurich
Try it out on
Earth System
FM Blog