COP-GEN

LATENT DIFFUSION TRANSFORMER

FOR COPERNICUS EARTH OBSERVATION DATA

Miguel Espinosa*[1], Eva Gmelich Meijling[2], Valerio Marsocci[2], Elliot J. Crowley[1], Mikolaj Czerkawski[3],

[1]University of Edinburgh, [2]European Space Agency (ESA), [3]Asterisk Labs

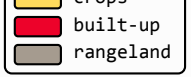
THE UNIVERSITY of EDINBURGH
School of Engineering

Φ-lab

Centre for Satellite Data in Environmental Science

DEM

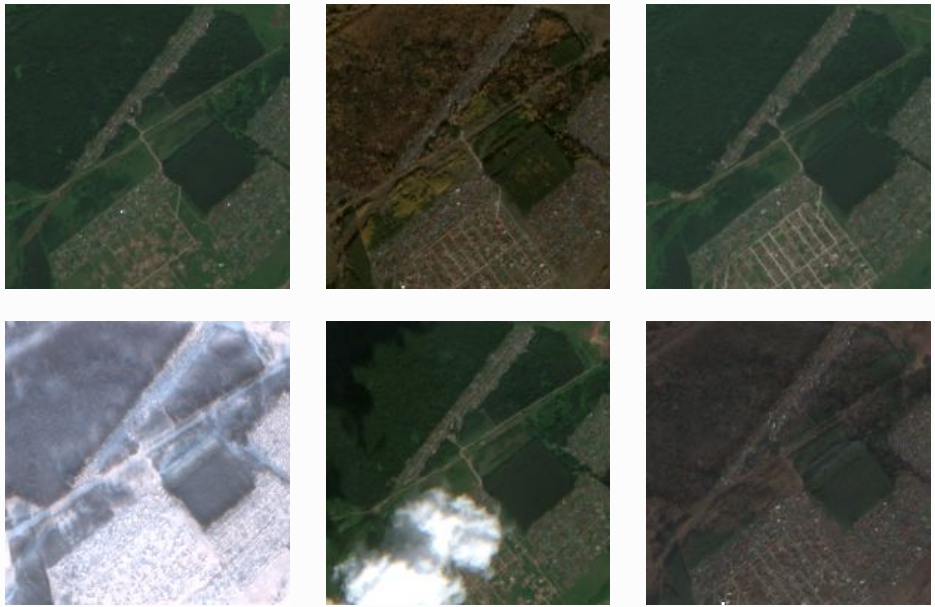
LULC



Which is the correct S2 image?

DEM

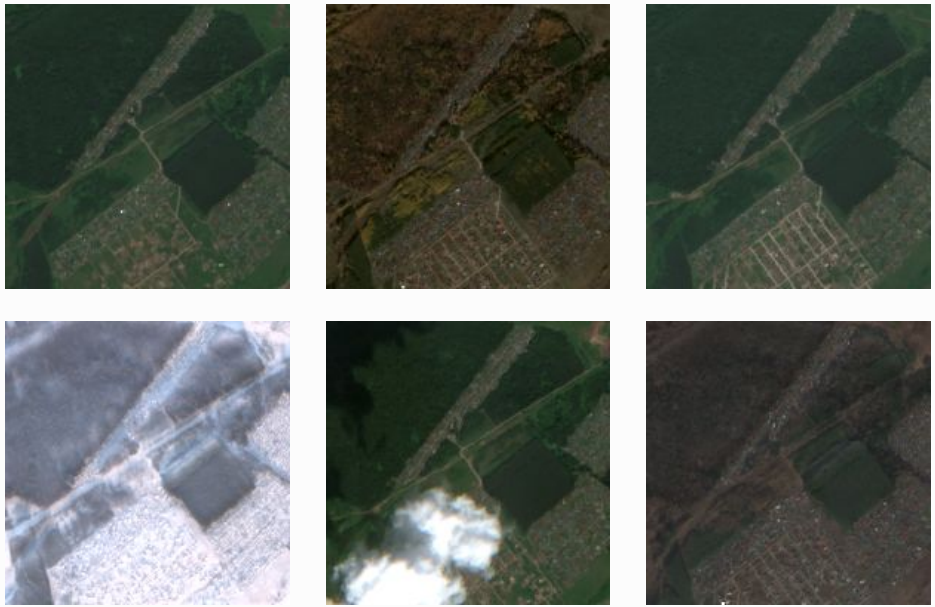
LULC



Which is the correct S2 image?

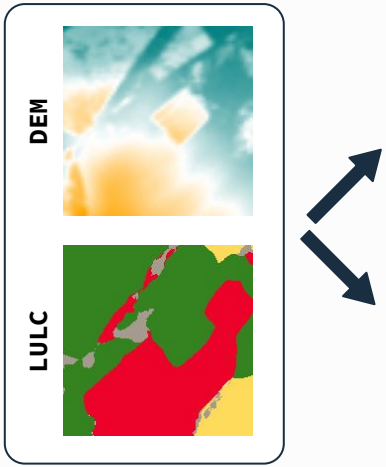
DEM

LULC



Cross-modal mappings in EO are non-injective

➔ yet... Generative FMs collapse towards the average values, failing to capture the distribution of possible outputs

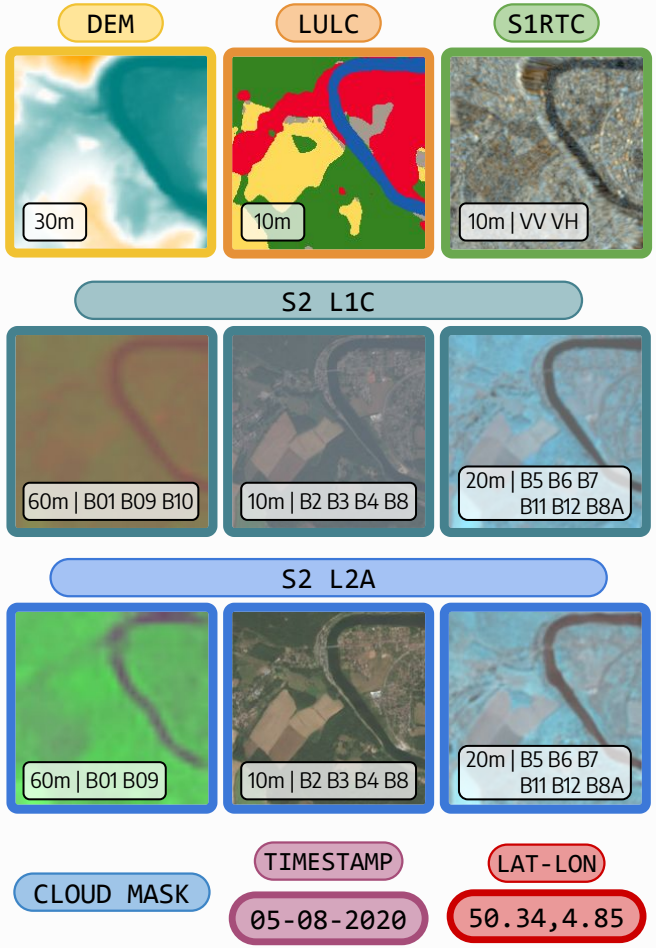


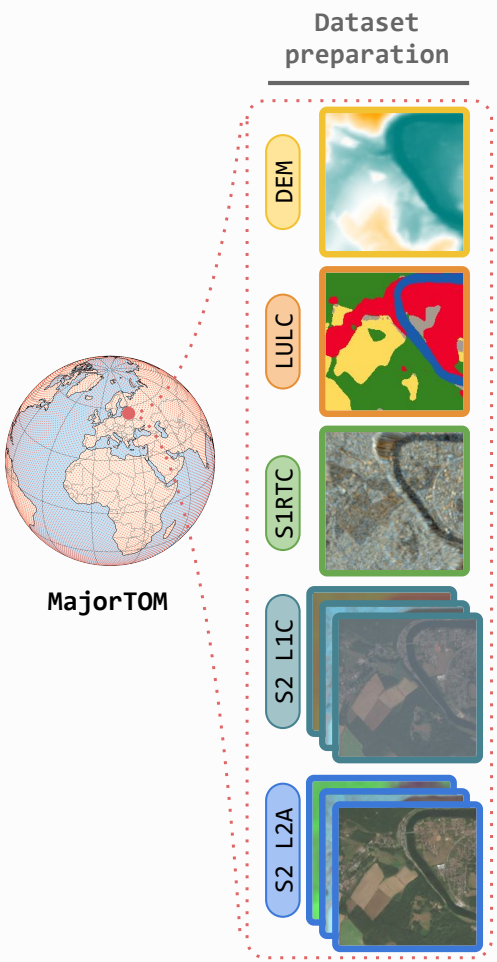
REAL S2L2A

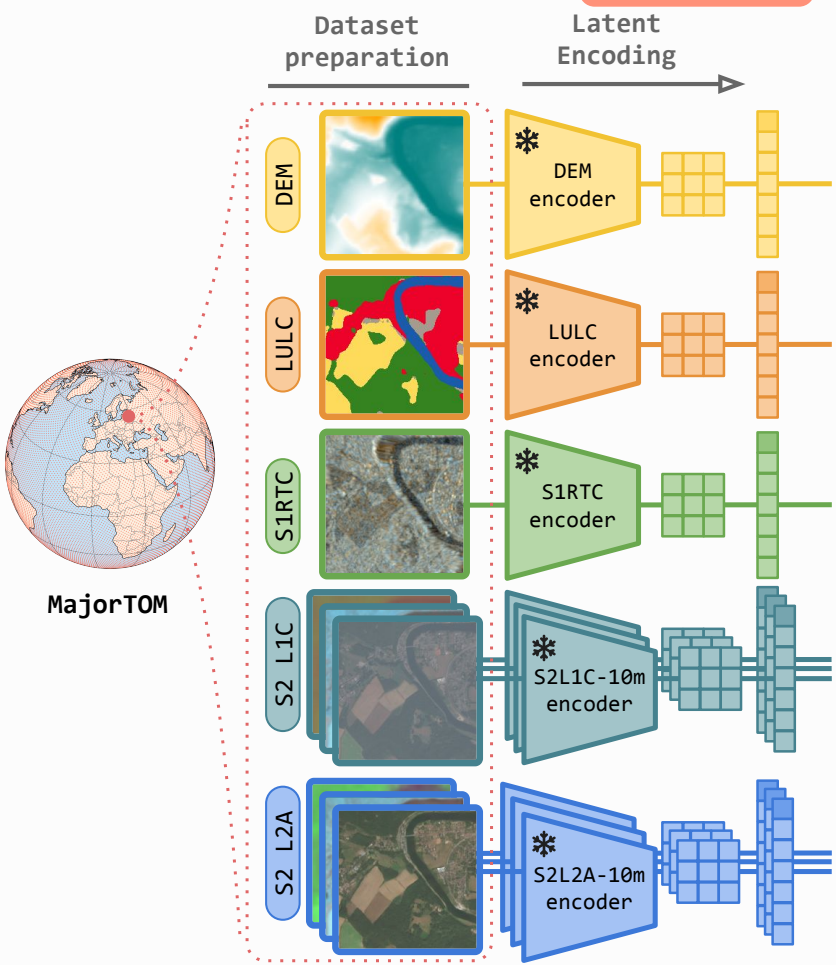
TERRAMIND

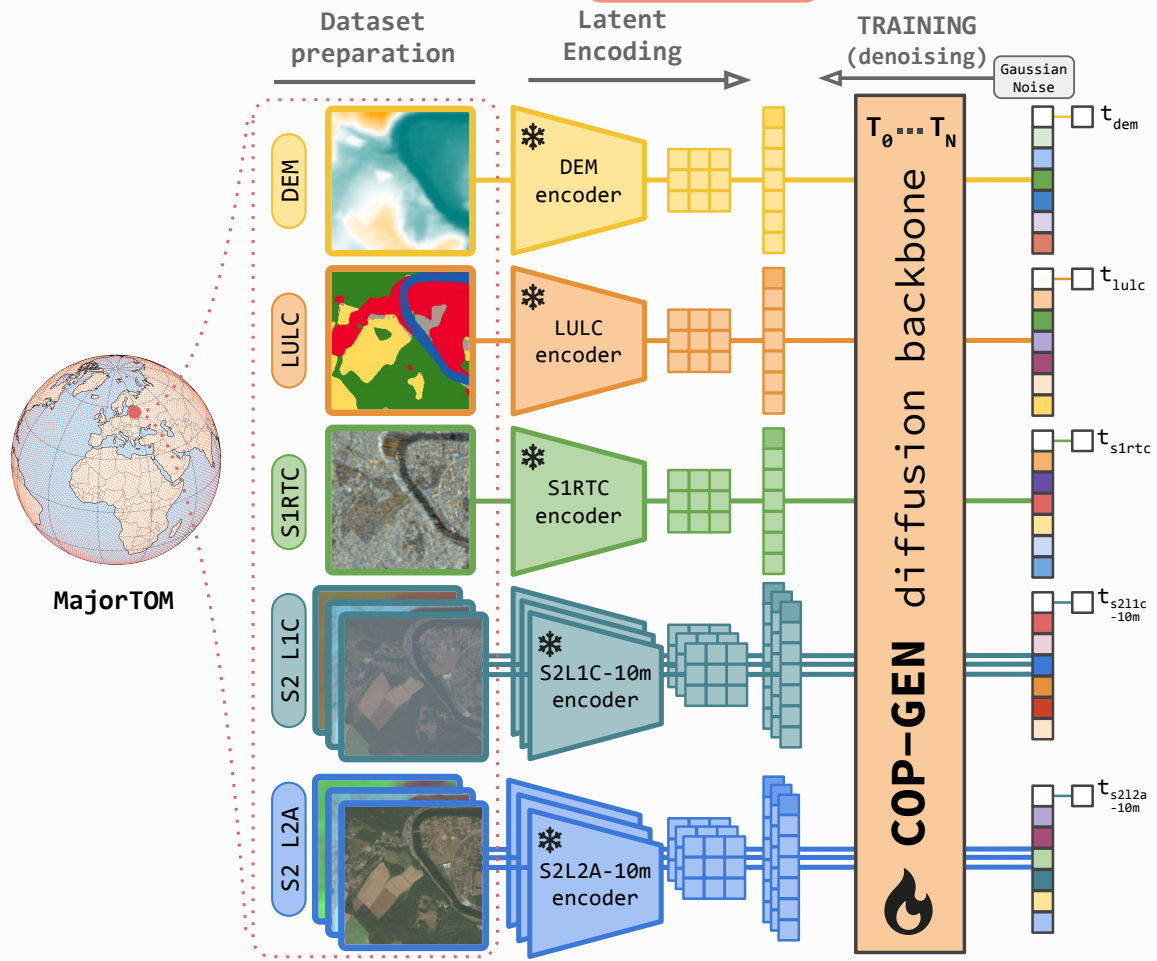
→ **COP-GEN**: Latent Diffusion Transformer for Copernicus EO Data

- Multi-modal inputs (native spatial-spectral resolution)
- Shared backbone (unified joint modelling)
- Any-to-any conditional generation (zero-shot)
- Output diversity (covering true distribution)




Dataset preparation
DEM
LULC
S1RTC
S2 L1C
S2 L2A
MajorTOM


Dataset preparation
Latent Encoding
MajorTOM
DEM
LULC
S1RTC
S2 L1C
S2 L2A
DEM encoder
LULC encoder
S1RTC encoder
S2L1C-10m encoder
S2L2A-10m encoder

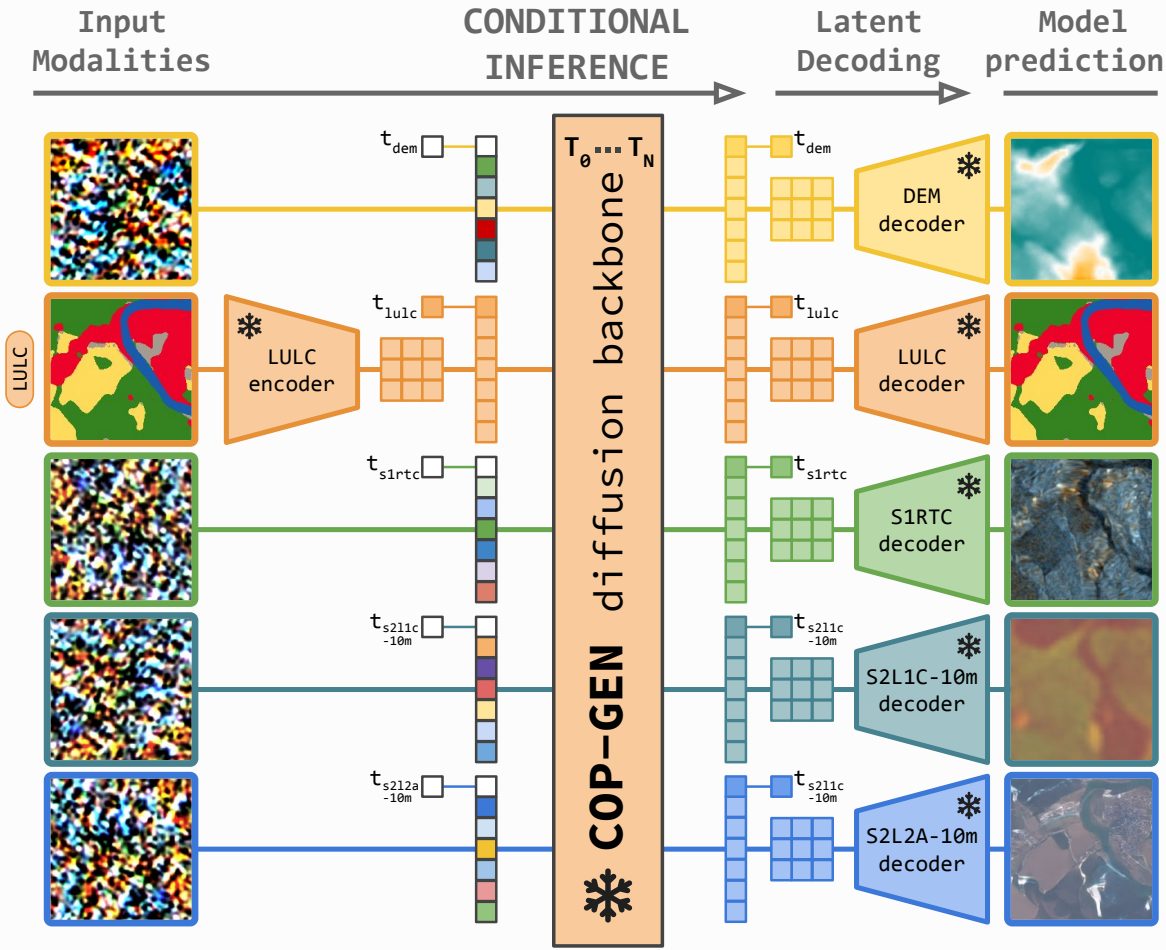


TRAINING:

- Independent noise levels for each modality
- Shared backbone promotes cross-modal information exchange

COP-GEN

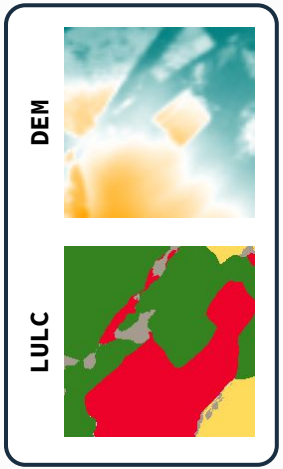
- U-ViT transformer architecture
- Streams of tokens



INFERENCE:

- Any-to-any conditional generation by simply adjusting the timesteps of the modalities separately

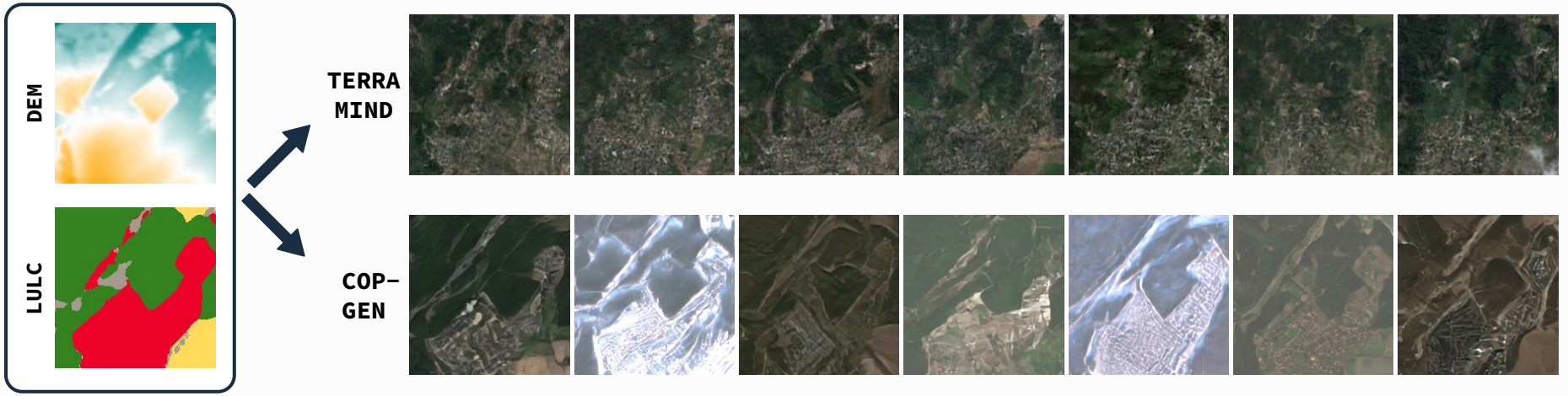
ANY-TO-ANY GENERATION:



REAL S2L2A

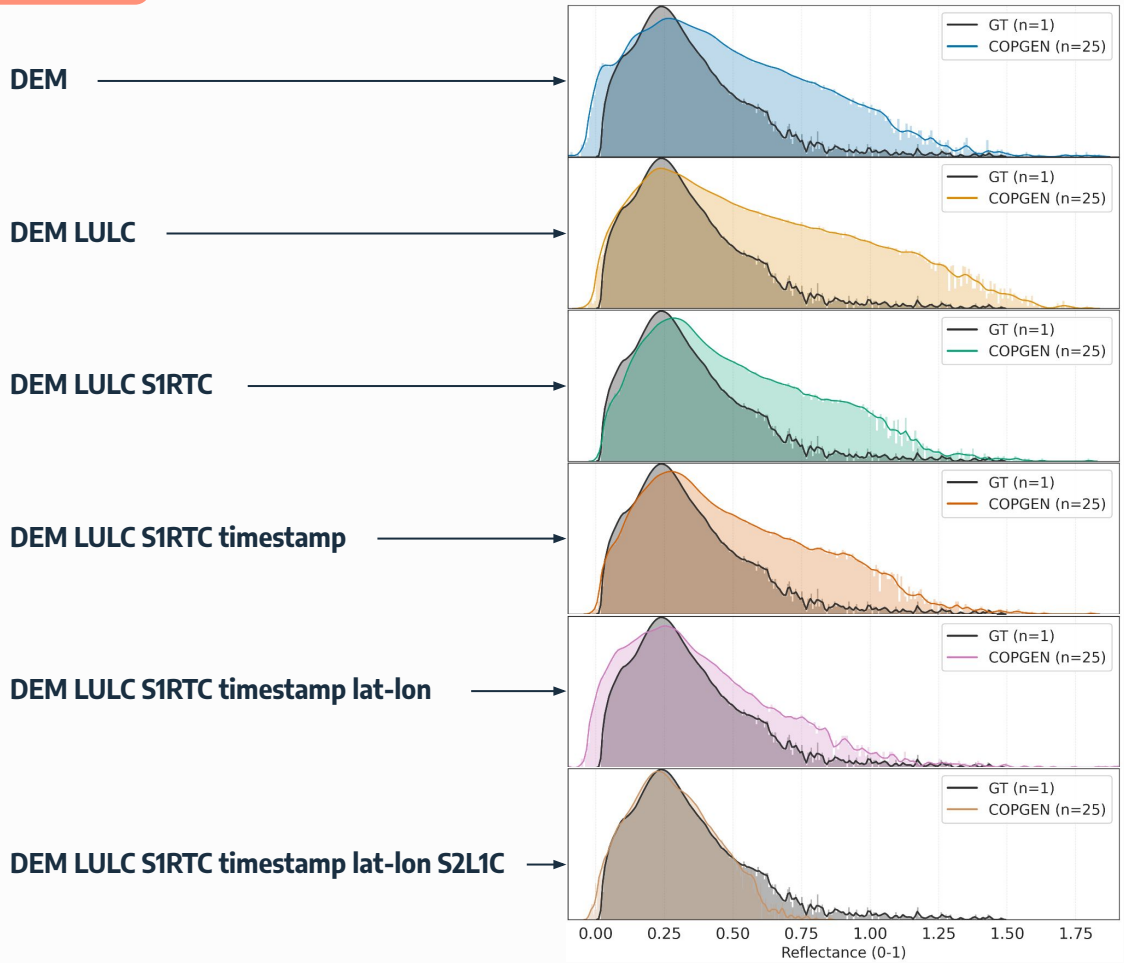
TERRA MIND

ANY-TO-ANY GENERATION:



Distribution spread narrowing by increasing input conditioning

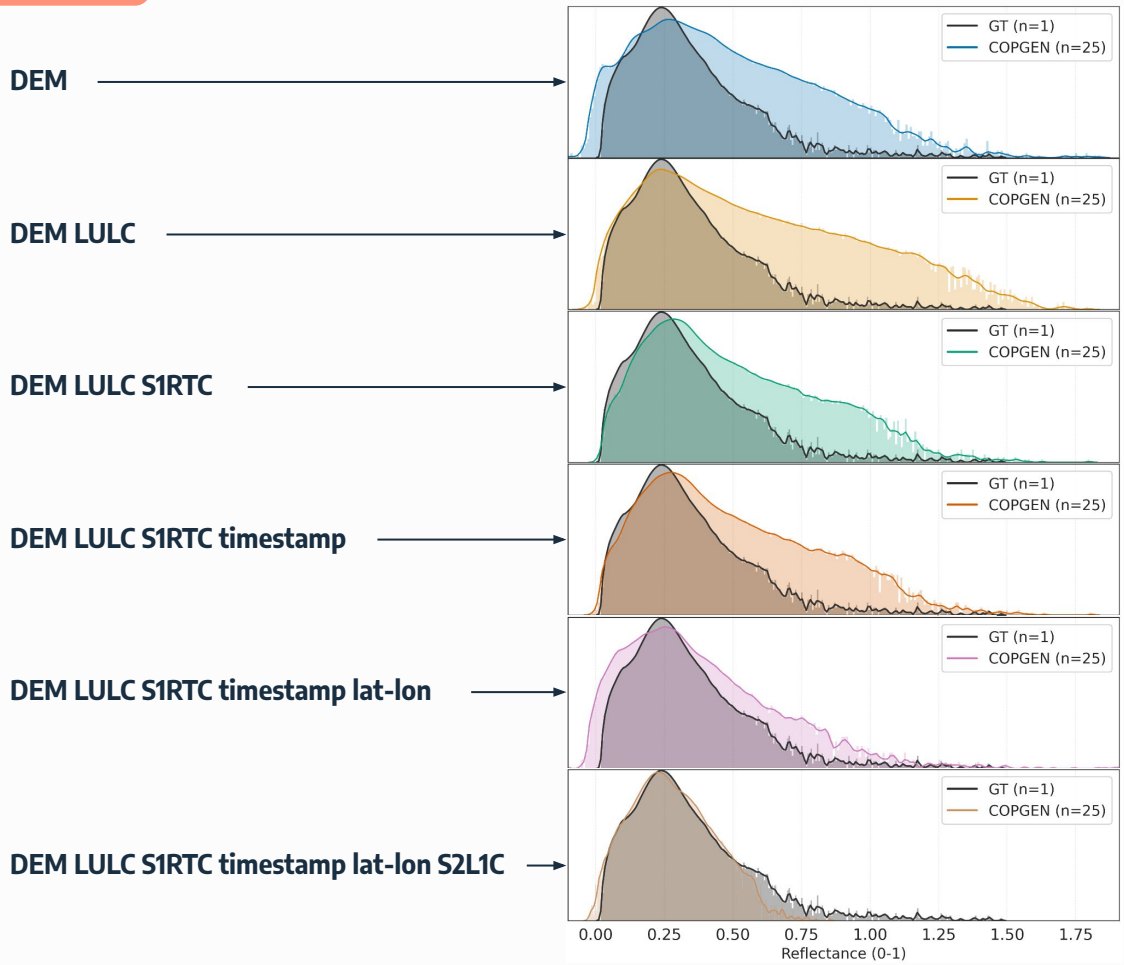
- ➔ Generate B07 - S2L2A
- ➔ Plot histogram of 25 COP-GEN generations
- ➔ *Bayesian estimator*



Distribution spread narrowing by increasing input conditioning

- ➔ Generate B07 - S2L2A
- ➔ Plot histogram of 25 COP-GEN generations
- ➔ *Bayesian estimator*

How do we evaluate *quantitatively* ?



COP-GEN Benchmark

*Does the model's output **distribution** match the true **distribution** of plausible observations?*

- ➔ ~500 locations
  - ◆ Per location:
    - 16 real S2L2A acquisitions
    - 16 Terramind S2L2A generations
    - 16 COP-GEN S2L2A generations
- ➔ New metrics: **evaluate distributions**, not samples
  - ◆ Perceptual fidelity metrics (embedding space)
  - ◆ Physical consistency metrics (spectral vectors)

COP-GEN Benchmark

*Does the model's output **distribution** match the true **distribution** of plausible observations?*

- ➔ ~500 locations
  - ◆ Per location:
    - 16 real S2L2A acquisitions
    - 16 Terramind S2L2A generations
    - 16 COP-GEN S2L2A generations
- ➔ New metrics: **evaluate distributions**, not samples
  - ◆ Perceptual fidelity metrics (embedding space)
  - ◆ Physical consistency metrics (spectral vectors)

	COP-GEN	Terramind
Recall	90.0%	2.8%

***COP-GEN** covers **90%** of the real observation manifold; TerraMind covers 2.8%.*

	COP-GEN	Terramind
Spectral coverage	63%	18%

***COP-GEN** spans **63%** of the real per-band reflectance range; TerraMind spans only 18%.*

Conclusions

- Mappings are not one-to-one → Model them as **distributions**

Conclusions

- Mappings are not one-to-one → Model them as **distributions**
- **COP-GEN** → the first to learn a joint generative distribution across modalities at native resolution, with any-to-any conditioning.

Conclusions

- Mappings are not one-to-one → Model them as **distributions**
- **COP-GEN** → the first to learn a joint generative distribution across modalities at native resolution, with any-to-any conditioning.
- **COP-GEN Benchmark** → Benchmark for evaluating distribution of outputs. Single-reference metrics hide diversity collapse.

Conclusions

- Mappings are not one-to-one → Model them as **distributions**
- **COP-GEN** → the first to learn a joint generative distribution across modalities at native resolution, with any-to-any conditioning.
- **COP-GEN Benchmark** → Benchmark for evaluating distribution of outputs. Single-reference metrics hide diversity collapse.

Code

Website

Paper

Benchmark