COP-GEN-Beta

Unified Generative Modelling of COPernicus Imagery Thumbnails

Miguel Espinosa*[1,2], Valerio Marsocci*[2], Yuru Jia[3], Elliot J. Crowley[1], Mikolaj Czerkawski[2,4]

[1]University of Edinburgh, [2]European Space Agency (ESA), [3]KU Leuven, [4]Asterisk Labs

github.com/miquel-espinosa/cop-gen-beta

*We introduce **COP-GEN-Beta**, a diffusion model designed to translate between remote sensing modalities.*

Motivation

- The alignment and integration of multiple sensor modalities is a prominent challenge in the field of Earth Observation
- There is a need for generative models that can:
  - capture the complex statistical relationships between multiple modalities
  - be used as a source of prior (instead of considering a single modality at a time)

With COP-GEN-Beta:

- *Flexibility*: Generate any combination, eliminating the need for specialized models
- *Unified multi-modal*: captures cross-modal relationships through a shared backbone, leveraging correlations between different data types
- *Scalability*: Adaptable framework ready to incorporate emerging remote sensing data types and modalities

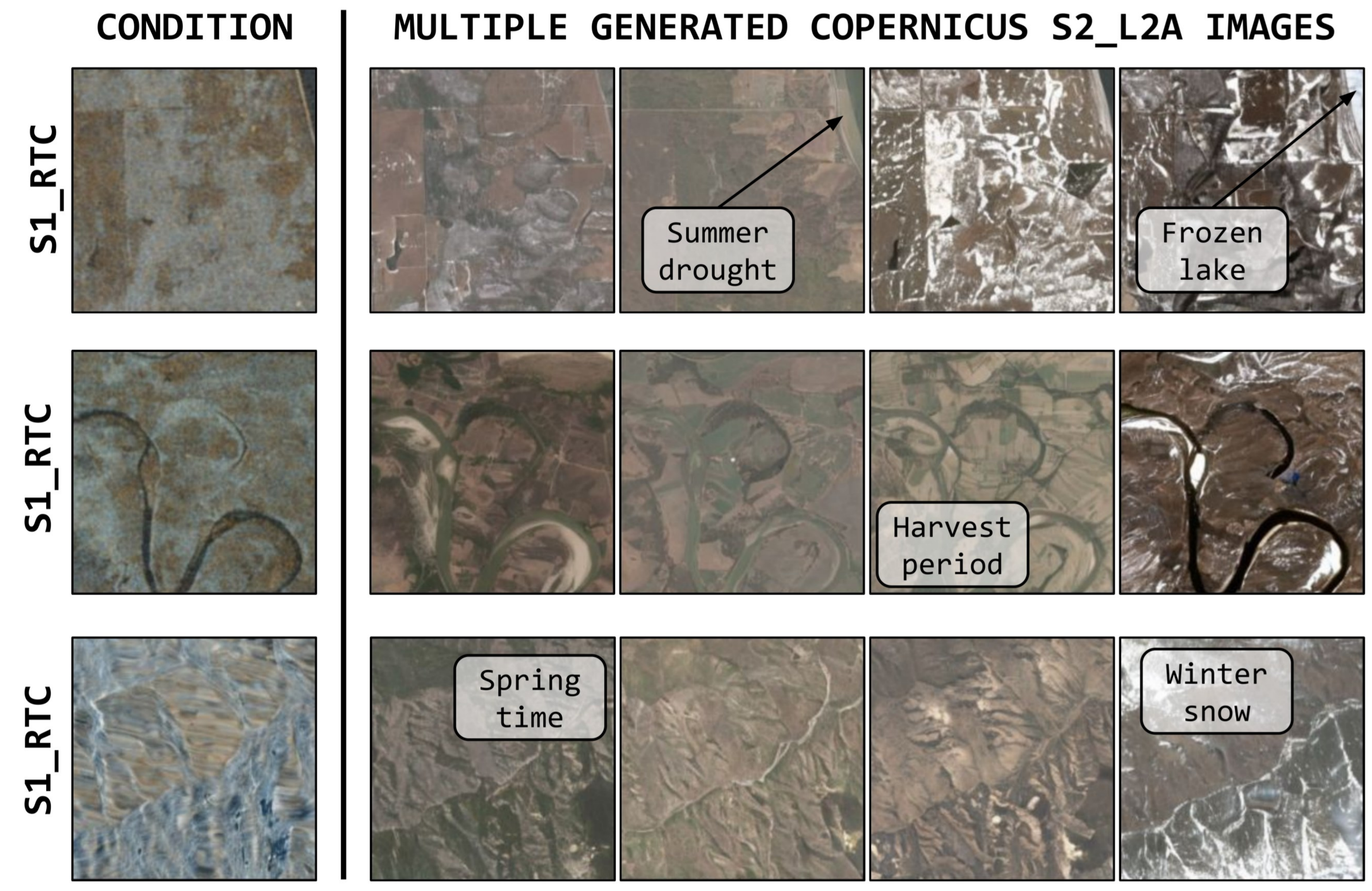

Fig 1. COP-GEN-Beta exhibits emergent effects, such as seasonality, when conditioned on the same S1RTC sample. It is capable of reimagining existing locations in conditions never observed.

Method

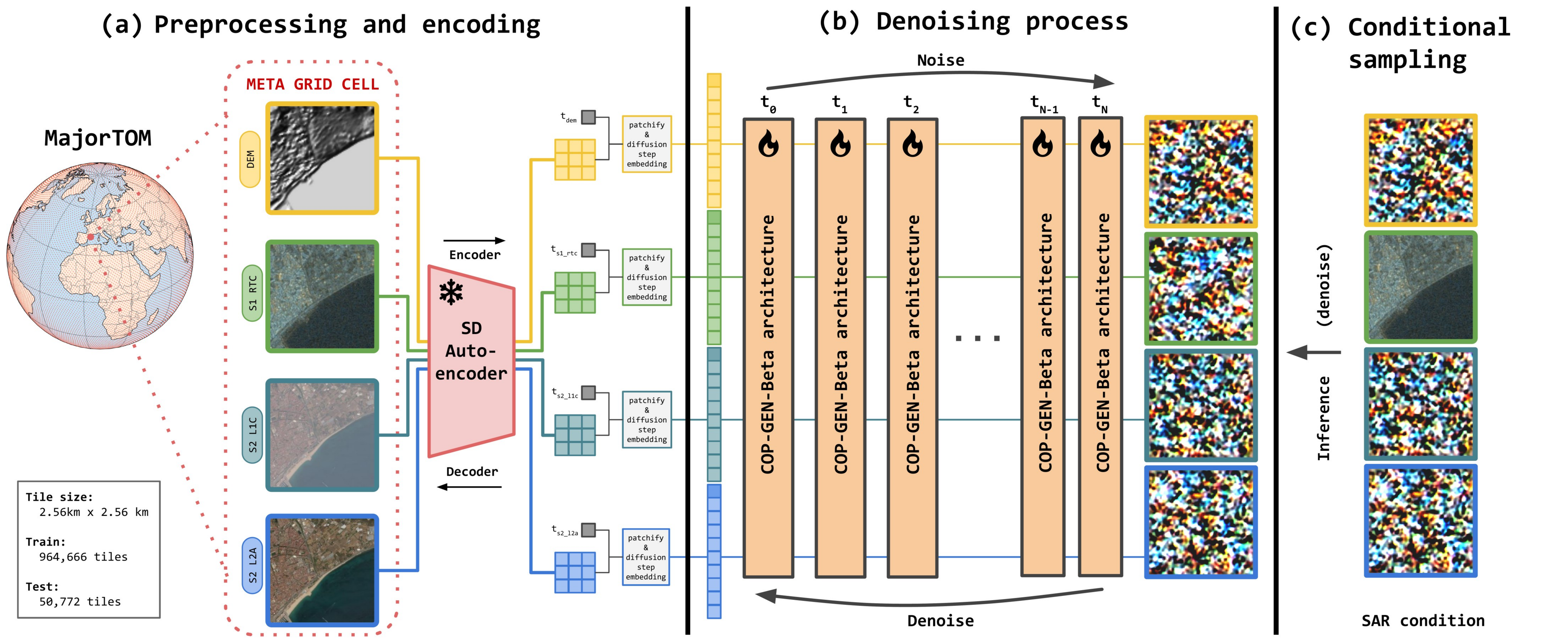

Fig 2a. Sample a global dataset of 4 modalities from Major TOM, encoding all images with SD autoencoder.

Fig 2b. Train a sequence-based diffusion model, where each modality has its own timestep.

Fig 2c. Generate all mod based on any subset.

- **Sequence-based** denoising diffusion model
  - Different data modalities are concatenated as latent token sequences
  - Having a shared backbone promotes cross-modal information exchange
  - Exhibits a scalable architecture where new modalities are added as extra input tokens.
- **Modality-specific timestep embeddings**
  - Provide full control of the generation
  - Adjustment of the diffusion timesteps per modality allows sampling different distributions at inference time

Results

- Outperforms DiffusionSAT in quantitative metrics (FID, Precision, Recall, F-Score)
- Strong qualitative results across diverse downstream tasks through flexible conditional and unconditional sampling
- Generalisation to unseen datasets (BigEarthNet)

Model	Condition	F-Score ↑ (%)				Precision ↑ (%)				Recall ↑ (%)				FID ↓			
		DEM	S1RTC	S2L1C	S2L2A	DEM	S1RTC	S2L1C	S2L2A	DEM	S1RTC	S2L1C	S2L2A	DEM	S1RTC	S2L1C	S2L2A
DiffusionSat [23]	text prompt	-	-	1.37	3.43	-	-	0.73	1.97	-	-	10.04	13.32	-	-	253.31	216.37
COP-GT	uncond.	42.25	24.22	36.71	39.54	73.25	**49.81**	73.65	71.72	29.69	16.00	24.45	27.29	90.36	52.21	53.66	48.49
COP-GT	DEM	NA	30.14	42.71	46.64	NA	45.35	68.04	67.77	NA	22.58	31.12	35.56	NA	23.97	27.33	25.32
COP-GT	S1RTC	49.17	NA	42.67	46.63	**65.53**	NA	68.96	69.38	39.35	NA	30.89	35.12	49.27	NA	29.55	27.18
COP-GT	S2L1C	**51.89**	32.57	NA	**83.51**	64.25	47.22	NA	**85.68**	43.52	24.86	NA	**81.45**	40.90	**20.94**	NA	**7.71**
COP-GT	S2L2A	51.41	**33.13**	**82.78**	NA	62.71	46.75	**85.50**	NA	**43.56**	**25.66**	**80.24**	NA	**36.70**	20.98	**5.57**	NA

Tab 1. Quantitative comparison. Conditioning on modalities improves performance, showing the advantages of guided gen.

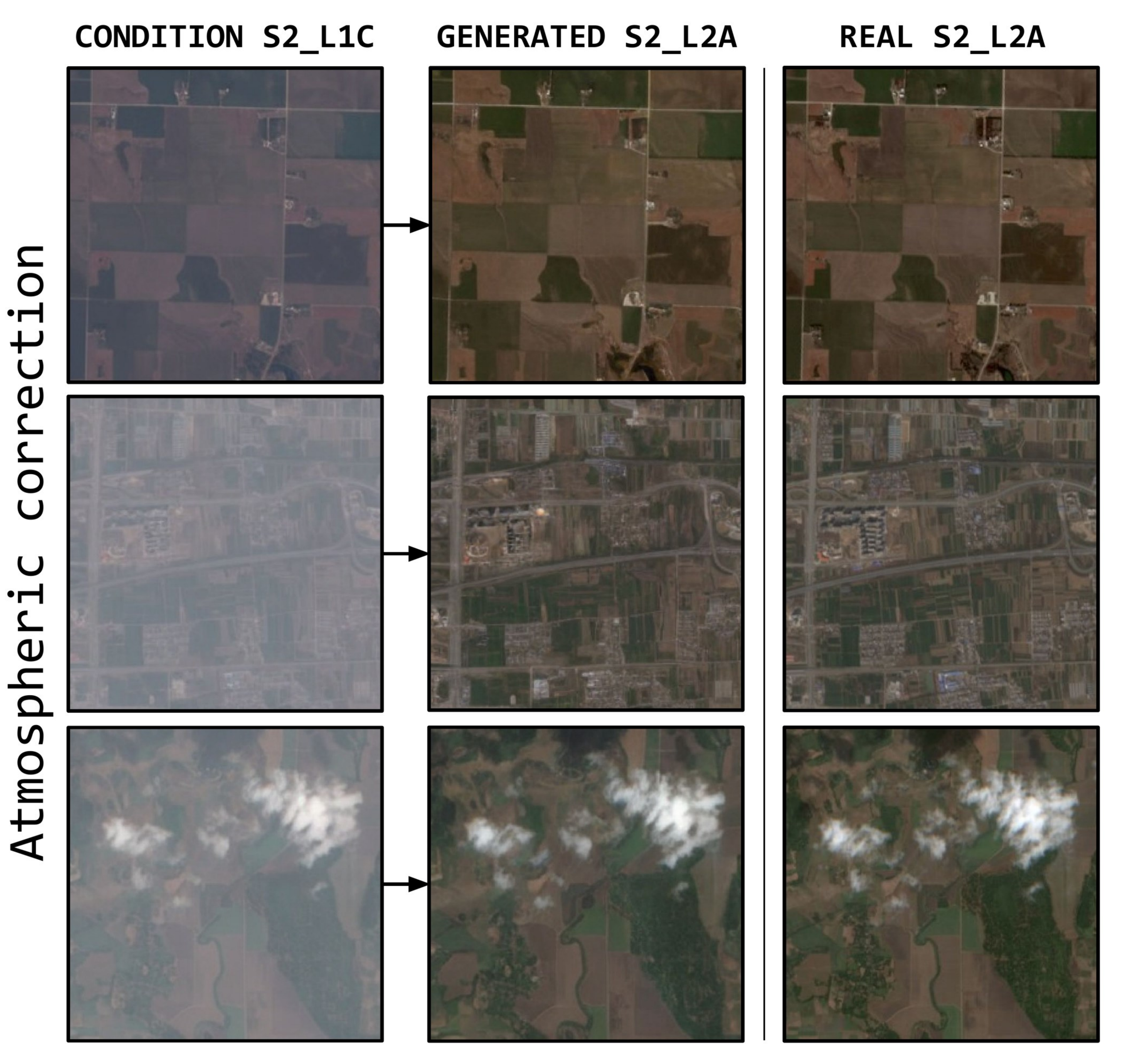

Fig 3a. Translate Top-of-Atmosphere input to Bottom-of Atmosphere, emulating the official processor for the L2A level.

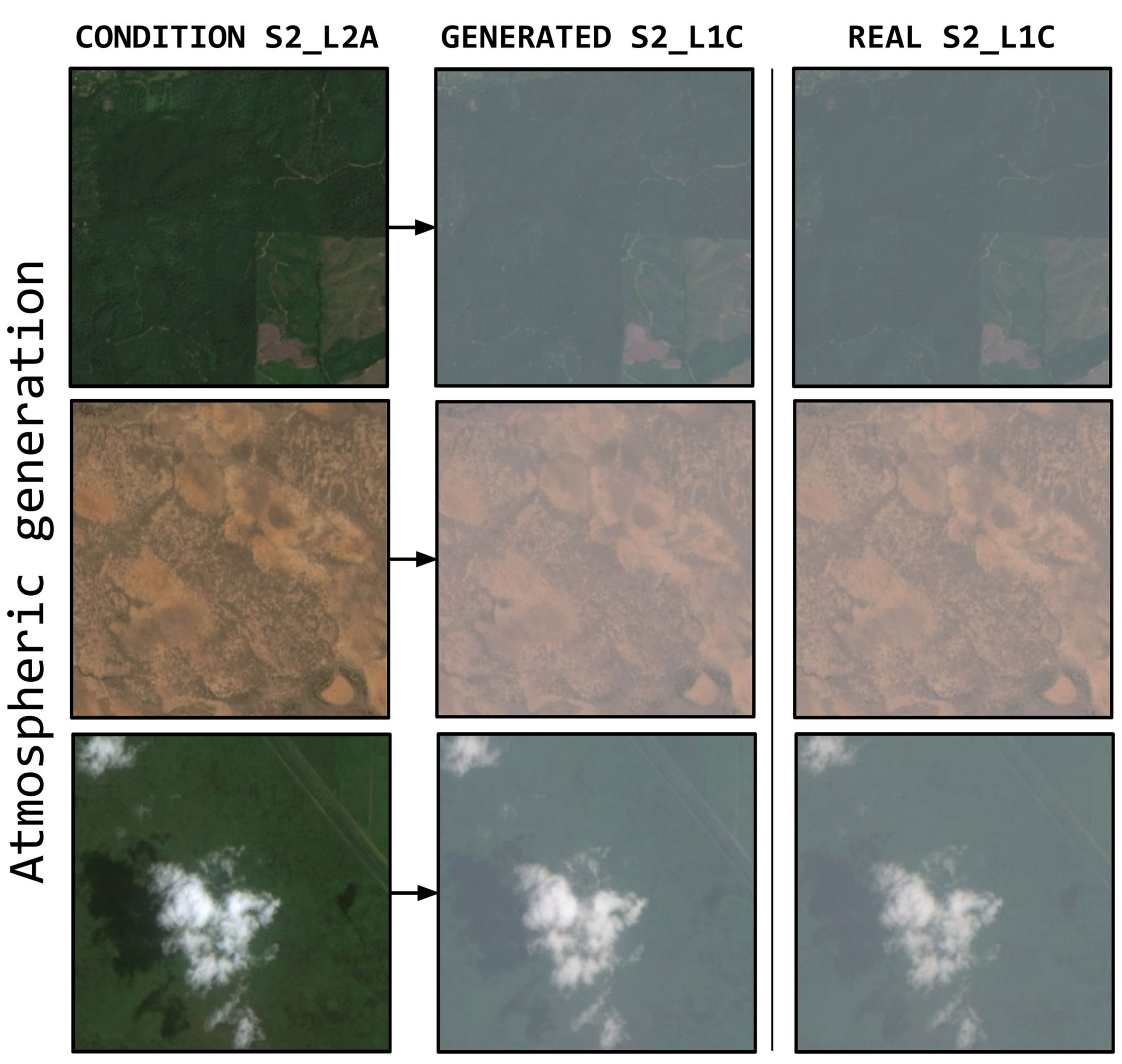

Fig 3b. Approximate a possible L1C product from an observed L2A observation, equivalent to the synthesis of the atmospheric effect.

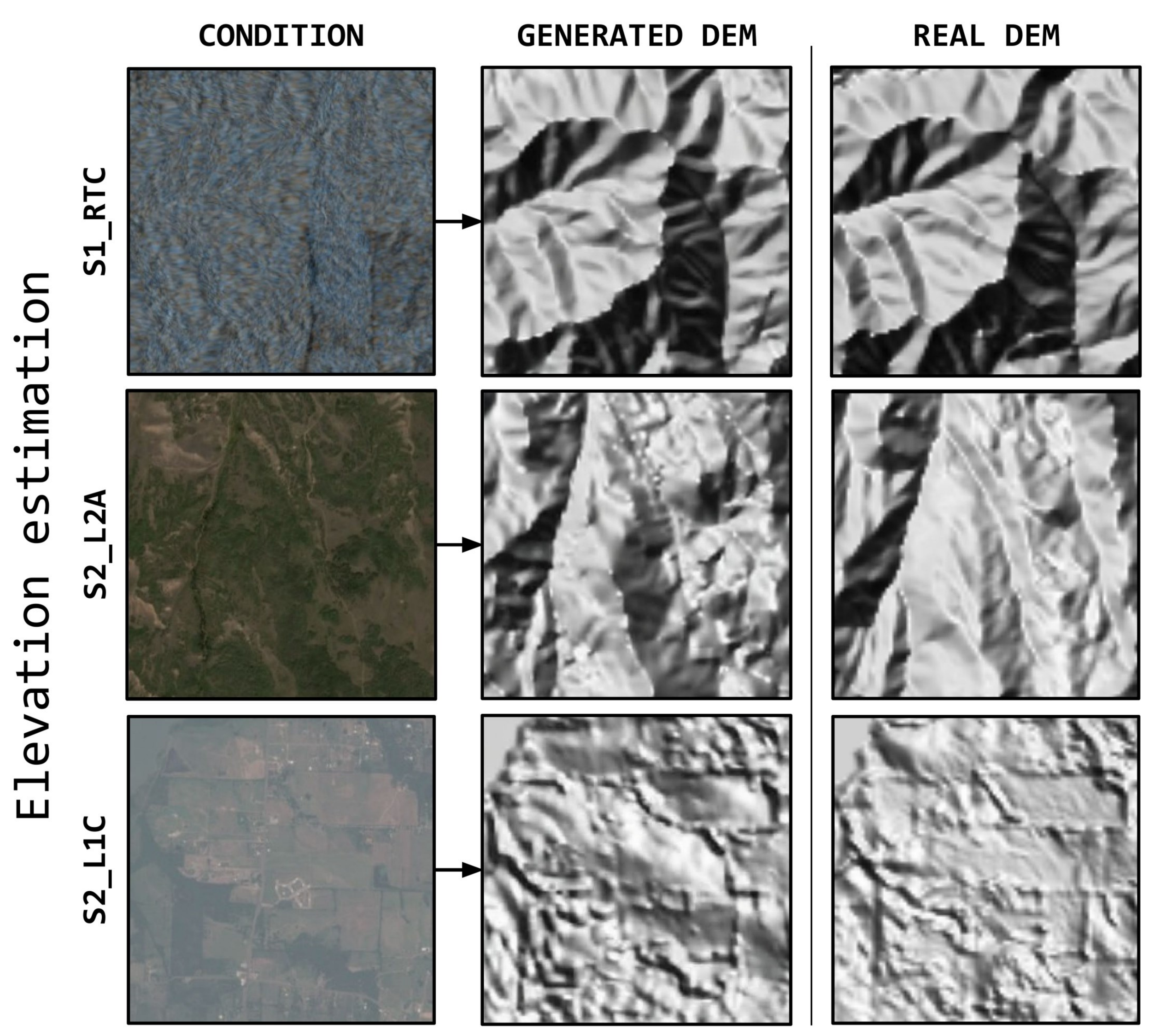

Fig 3c. Map any modality to an elevation model estimate (highly useful for dynamically changing terrains).

Acknowledgements

Funding for this research is provided by a studentship from the School of Engineering at the University of Edinburgh, the SENSE - Centre for Satellite Data in Environmental Science CDT